(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

PCT

(43) International Publication Date
26 May 2005 (26.05.2005)

(10) International Publication Number
WO 2005/047523 A2

(51) International Patent Classification⁷: C12Q

(21) International Application Number:
PCT/US2004/037613

(22) International Filing Date:
12 November 2004 (12.11.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/519,862	12 November 2003 (12.11.2003)	US
60/546,277	19 February 2004 (19.02.2004)	US
60/547,611	24 February 2004 (24.02.2004)	US

(71) Applicant *(for all designated States except US)*: HELI-COS BIOSCIENCES CORPORATION [US/US]; One Kendall Square, Building 200, Cambridge, MA 02139 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LAPIDUS, Stanley, N. [US/US]; 12 Old Evergreen Road, Beford, NH 03110 (US). BUZBY, Phillip, Richard [US/US]; 31 Harlan Drive, Andover, MA 02301-5878 (US). HARRIS, Timothy, A. [US/US]; 1387 Church Street, Ocean County, NJ 08755 (US).

(74) Agent: CHUNG, Christopher, H.; Proskauer Rose LLP, One International Place 14th Floor, Boston, MA 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SHORT CYCLE METHODS FOR SEQUENCING POLYNUCLEOTIDES

(57) Abstract: The invention provides methods for sequencing a polynucleotide comprising stopping an extension cycle in a sequence by synthesis reaction before the reaction has run to near of full completion.

Short Cycle Methods For Sequencing Polynucleotides

Related Applications

[0001]   This application claims the benefit of U.S. Provisional Application Nos. 60/546,277, filed on February 19, 2004, 60/547,611, filed on February 24, 2004, and 60/519,862, filed on November 12, 2003.

Field of the Invention

[0002]   The invention relates to methods for sequencing a polynucleotide, and more particularly, to methods for high throughput single molecule sequencing of target polynucleotides.

Background

[0003] .   Completion of the human genome has paved the way for important insights into biologic structure and function. Knowledge of the human genome has given rise to inquiry into individual differences, as well as differences within an individual, as the basis for differences in biological function and dysfunction. For example, single nucleotide differences between individuals, called single nucleotide polymorphisms (SNPs), are responsible for dramatic phenotypic differences. Those differences can be outward expressions of phenotype or can involve the likelihood that an individual will get a specific disease or how that individual will respond to treatment. Moreover, subtle genomic changes have been shown to be responsible for the manifestation of genetic diseases, such as cancer. A true understanding of the complexities in either normal or abnormal function will require large amounts of specific sequence information.

[0004]   An understanding of cancer also requires an understanding of genomic sequence complexity. Cancer is a disease that is rooted in heterogeneous genomic instability. Most cancers develop from a series of genomic changes, some subtle and some significant, that occur in a small subpopulation of cells. Knowledge of the sequence variations that lead to cancer will lead to an understanding of the etiology of the disease, as well as ways to treat and prevent it. An essential first step in understanding genomic complexity is the ability to perform high-resolution sequencing.

[0005]   Various approaches to nucleic acid sequencing exist. One conventional way to do bulk sequencing is by chain termination and gel separation, essentially as described by Sanger et al., Proc Natl Acad Sci U S A, 74(12): 5463-67 (1977). That method relies on the generation of a mixed population of nucleic acid fragments representing terminations at each

base in a sequence. The fragments are then run on an electrophoretic gel and the sequence is revealed by the order of fragments in the gel. Another conventional bulk sequencing method relies on chemical degradation of nucleic acid fragments. See, Maxam et al., Proc. Natl. Acad. Sci., 74: 560-564 (1977). Finally, methods have been developed based upon

5   sequencing by hybridization. See, e.g., Drmanac, et al., Nature Biotech., 16: 54-58 (1998). Bulk techniques, such as those described above, cannot effectively detect single nucleotide differences between samples, and are not useful for comparative whole genome sequencing. Single molecule techniques are necessary for high-resolution detection of sequence differences.

10   [0006]   There have been several recent reports of sequencing using single molecule techniques. Most conventional techniques have proposed incorporation of fluorescently-labeled nucleotides in a template-dependent manner. A fundamental problem with conventional single molecule techniques is that the sequencing reactions are run to completion. For purposes of single molecule chemistry, this typically means that template is

15   exposed to nucleotides for incorporation for about 10 half lives. This gives rise to problems in the ability to resolve single nucleotides as they incorporate in the growing primer strand. The resolution problem becomes extreme in the situation in which the template comprises a homopolymer region. Such a region is a continuous sequence consisting of the same nucleotide species. When optical signaling is used as the detection means, conventional

20   optics are able to reliably distinguish one from two identical bases, and sometimes two from three, but rarely more than three. Thus, single molecule sequencing using fluorescent labels in a homopolymer region typically results in a signal that does not allow accurate determination of the number of bases in the region.

[0007]   One method that has been developed in order to address the homopolymer

25   issue provides for the use of nucleotide analogues that have a modification at the 3' carbon of the sugar that reversibly blocks the hydroxyl group at that position. The added nucleotide is detected by virtue of a label that has been incorporated into the 3' blocking group. Following detection, the blocking group is cleaved, typically, by photochemical means to expose a free hydroxyl group that is available for base addition during the next cycle.

30   [0008]   However, techniques utilizing 3' blocking are prone to errors and inefficiencies. For example, those methods require excessive reagents, including numerous primers complementary to at least a portion of the target nucleic acids and differentially-labeled nucleotide analogues. They also require additional steps, such as cleaving the

- 3 -

blocking group and differentiating between the various nucleotide analogues incorporated into the primer. As such, those methods have only limited usefulness.

[0009]   Need therefore exists for more effective and efficient methods and devices for single molecule nucleic acid sequencing.

Summary of the Invention

[0010]   The invention provides methods for high throughput single molecule sequencing. In particular, the invention provides methods for controlling at least one parameter of a nucleotide extension reaction in order to regulate the rate at which nucleotides are added to a primer. The invention provides several ways of controlling nucleic acid sequence-by-synthesis reactions in order to increase the resolution and reliability of single molecule sequencing. Methods of the invention solve the problems that imaging systems have in accurately resolving a sequence at the single-molecule level. In particular, methods of the invention solve the problem of determining the number of nucleotides in a homopolymer stretch.

[0011]   Methods of the invention generally contemplate terminating sequence-by-synthesis reactions prior to completion in order to obtain increased resolution of individual nucleotides in a sequence. Fundamentally, this requires exposing nucleotides to a mixture comprising a template, a primer, and a polymerase under conditions sufficient for only limited primer extension. Reactions are conducted under conditions such that it is statistically unlikely that more than 1 or 2 nucleotides are added to a growing primer strand in any given incorporation cycle. An incorporation cycle comprises exposure of a template/primer to nucleotides directed at the base immediately downstream of the primer (this may be all four conventional nucleotides or analogs if the base is not known) and washing unhybridized nucleotide.

[0012]   Nucleotide addition in a sequence-by-synthesis reaction is a stochastic process. As in any chemical reaction, nucleotide addition obeys the laws of probability. Methods of the invention are concerned with controlling the rate of nucleotide addition on a per-cycle basis. That is, the invention teaches ways to control the rate of nucleotide addition within an extension cycle given the stochastic nature of the extension reaction itself. Methods of the invention are intended to control reaction rates within the variance that is inherent in a reaction that is fundamentally stochastic. Thus, the ability to control, according to the invention, base addition reactions such that, on average, no more than two bases are added in any cycle takes into account the inherent statistics of the reactions.

[0013]   The invention thus teaches polynucleotide sequence analysis using short cycle chemistry. One embodiment of the invention provides methods for slowing or reversibly inhibiting the activity of polymerase during a sequencing-by-synthesis reaction. Other methods teach altering the time of exposure of nucleotides to the template-primer complex.

5   Still other methods teach the use of physical blockers that temporarily halt or slow polymerase activity and/or nucleotide addition. In general, any component of the reaction that permits regulation of the number of labeled nucleotides added to the primer per cycle, or the rate at which the nucleotides are incorporated and detected per cycle is useful in methods of the invention. Additional components include, but are not limited to, the presence or absence of a

10 label on a nucleotide, the type of label and manner of attaching the label; the linker identity and length used to attach the label; the type of nucleotide (including, for example, whether such nucleotide is a dATP, dCTP, dTTP, dGTP or dUTP; a natural or non-natural nucleotide, a nucleotide analogue, or a modified nucleotide); the "half-life" of the extension cycle (where one half-life is the time taken for at least one incorporation to occur in 50% of the

15 complementary strands); the local sequence immediately 3' to the addition position; whether such base is the first, second, third, etc. base added; the type of polymerase used; the particular batch characteristics of the polymerase; the processivity of the polymerase; the incorporation rate of the polymerase; the number of wash cycles (i.e., the number of times a nucleotide is introduced to the reaction then washed out); the number of target nucleic acids in

20 the reaction; the temperature of the reaction and the reagents used in the reaction.

[0014]   In a preferred embodiment of the invention, a nucleic acid template is exposed to a primer capable of hybridizing to the template and a polymerase capable of catalyzing nucleotide addition to the primer. A labeled nucleotide is introduced for a period of time that is statistically insufficient for incorporation of more than about 2 nucleotides per cycle.

25 Nucleotide exposure may also be coordinated with polymerization inhibition such that, on average, 0, 1, or 2 labeled nucleotides are added to the primer, but that 3 labeled nucleotides are almost never added to the primer in each cycle. Ideally, the exposure time, during which labeled nucleotides are exposed to the template-primer complex, is statistically insufficient for incorporation of more nucleotides than are resolvable by a detection system used to detect

30 incorporation.

[0015]   The invention also contemplates performing a plurality of base incorporation cycles. Each cycle comprises exposing a template nucleic acid to a labeled nucleotide that is not a chain-terminating nucleotide. The labeled nucleotide is incorporated into a primer

hybridized to the template nucleic acid if the nucleotide is capable of hybridizing to the template nucleotide immediately upstream of the primer and there is about a 99% probability that two or fewer of said nucleotides are incorporated into the same primer strand per cycle. Incorporated nucleotides are then identified.

5   [0016]   Methods of the invention also make use of differential base incorporation rates in order to control overall reaction rates. For example, the rate of incorporation is lower for a second nucleotide given incorporation of a prior nucleotide immediately upstream of the second. This effect is magnified if the first nucleotide comprises a label or other group that hinders processivity of the polymerase. By determining an approximate reduction in the rate

10   of incorporation of the second nucleotide, one can regulated the time of exposure of a sample to a second labeled nucleotide such that the time is statistically insufficient for incorporation of more nucleotides than are resolvable by a detection system used to detect incorporation of the nucleotide into the primer.

[0017]   The invention may also be conducted using a plurality of primer extension

15   cycles, wherein each cycle comprises exposing a target nucleic acid to a primer capable of hybridizing to the target, thereby forming a primed target; exposing the primed target to a labeled nucleic acid in the presence of a nucleic acid polymerase, coordinating transient inhibition of the polymerase and time of exposure to the labeled nucleotide such that it is statistically likely that at least one of said labeled nucleic acid is incorporated in the primer,

20   but statistically unlikely that more than two of the labeled nucleotide are incorporated in the primer.

[0018]   According to another embodiment, methods of the invention comprise conducting a cycle of template-dependent nucleic acid primer extension in the presence of a polymerase and a labeled nucleotide; inhibiting polymerase activity such that it is statistically

25   unlikely that more than about 2 nucleotides are incorporated into the same primer strand in the cycle; washing unincorporated labeled nucleotide away from the template; detecting any incorporation of the labeled nucleotide; neutralizing label in any incorporated labeled nucleotide; removing the inhibition; repeating the foregoing steps; and compiling a sequence based upon the sequence of nucleotides incorporated into the primer.

30   [0019]   In another embodiment, the invention provides a method comprising exposing a nucleic acid template to a primer capable of hybridizing to a portion of the template in order to form a template/primer complex reaction mixture; adding a labeled nucleotide in the presence of a polymerase to the mixture under conditions that promote incorporation of the

nucleotide into the primer if the nucleotide is complementary to a nucleotide in the template that is downstream of said primer; coordinating removal of the labeled nucleotide and inhibition of the polymerase so that no more than about 2 nucleotides are incorporated into the same primer; identifying labeled nucleotide that has been incorporated into said primer;

5   repeating the foregoing steps at least once; and determining a sequence of the template based upon the order of the nucleotides incorporated into the primer.

[0020]   According to another embodiment, the method comprises exposing a template nucleic acid to a primer capable of hybridizing to a portion of the template upstream of a region of the template to be sequenced; introducing a labeled nucleic acid and a polymerase to

10   the template under conditions wherein the labeled nucleic acid will be incorporated in the primer if the labeled nucleic acid is capable of hybridizing with a base downstream of the primer; and controlling the rate of the incorporation by limiting the time of exposure of the labeled nucleic acid to the template or by inhibiting the polymerase at a predefined time after exposure of the template to the labeled nucleotide; detecting incorporation of the labeled

15   nucleotide into the primer; and identifying the nucleotide in the template as the complement of labeled nucleotide incorporated into the primer.

[0021]   In yet another embodiment, methods of the invention comprise exposing a target polynucleotide to a primer capable of hybridizing to the polynucleotide, extending the primer in the presence of a polymerizing agent and one or more extendible nucleotides, each

20   comprising a detectable label. The polymerizing agent is exposed to a cofactor (i.e., any agent that decreases or halts polymerase activity), and the incorporation of label is detected. The steps of extending the primer and exposing the polymerizing agent to a cofactor may be performed simultaneously, or may be performed in separate steps. In one embodiment, the method further comprises inactivating the cofactor, thereby reversing its effect on the

25   polymerizing agent. Modes of inactivation depend on the cofactor. For example, where the cofactor is attached to the nucleotide, inactivation can typically be achieved by cleaving the cofactor from the nucleotide.

[0022]   Methods of the invention also address the problem of reduced detection due to a failure of some strands in a given cycle to incorporate labeled nucleotide. In each

30   incorporation cycle, a certain number of strands fail to incorporate a nucleotide that should be incorporated based upon its ability to hybridize to a nucleotide present in the template. The strands that fail to incorporate a nucleotide in a cycle will not be prepared to incorporate a nucleotide in the next cycle (unless it happens to be the same as the unincorporated

nucleotide, in which case the strand will still lag behind unless both nucleotides are incorporated in the same cycle). Essentially, this situation results in the strands that failed to incorporate being unavailable for subsequent polymerase-catalyzed additions to the primer. That, in turn, leads to fewer strands available for base addition in each successive cycle

5   (assuming the non-incorporation occurs in all or most cycles). The invention overcomes this problem by exposing a template/primer complex to a labeled nucleotide that is capable of hybridizing to the template nucleotide immediately downstream of the primer. After removing unbound labeled nucleotide, the sample is exposed to unlabeled nucleotide, preferably in excess, of the same species. The unlabeled nucleotide "fills in" the positions in

10   which hybridization of the labeled nucleotide did not occur. That functions to increase the number of strands that are available for participation in the next round. The effect is to increase resolution in subsequent rounds over background. In a preferred embodiment, the labeled nucleotide comprises a label that impedes the ability of polymerase to add a downstream nucleotide, thus temporarily halting the synthesis reaction until unlabeled

15   nucleotide can be added, at which point polymerase inhibition is removed and t he next incorporation cycle is conducted

[0023]   One feature of this embodiment is that a sequence is compiled based upon the incorporation data, while allowing maximum strand participation in each cycle. Thus, methods of the invention are useful for identifying placeholders in some strands in a

20   population of strands being sequenced. As long as there are no more than two consecutive placeholders in any one strand, the invention has a high tolerance for placeholders with little or no effect on the ultimate sequence determination.

[0024]   Methods of the invention are also useful for identifying a single nucleotide in a nucleic acid sequence. The method comprises the steps of sequentially exposing a template-

25   bound primer to a labeled nucleotide and an unlabeled nucleotide of the same type in the presence of a polymerase under conditions that allow template-dependent primer extension; determining whether the first nucleotide is incorporated in the primer at a first position; repeating the sequentially exposing step using subsequent labeled and unlabeled nucleotides until a nucleotide is identified at the first position.

30   [0025]   Identification of nucleotides in a sequence can be accomplished according to the invention using fluorescence resonance energy transfer (FRET). Single pair FRET (spFRET) is a good mechanism for increasing signal-to-noise in single molecule sequencing. Generally, a FRET donor (e.g., cyanine-3) is placed on the primer, on the polymerase, or on a

previously incorporated nucleotide. The primer/template complex then is exposed to a nucleotide comprising a FRET acceptor (e.g., cyanine-5). If the nucleotide is incorporated, the acceptor is activated and emits detectable radiation, while the donor goes dark. That is the indication that a nucleotide has been incorporated. The nucleotide is identified based upon

5   knowledge of which nucleotide species contained the acceptor. The invention also provides methods for identifying a placeholder in a nucleic acid sequence using FRET. A nucleic acid primer is hybridized to a target nucleic acid at a primer binding site in the target. The primer comprises a donor fluorophore. The hybridized primer is exposed to a first nucleotide comprising an acceptor fluorophore that, when incorporated into the primer, prevents further

10   polymerization of the primer. Whether there is fluorescent emission from the donor and the acceptor is determined, and a placeholder in the nucleic acid sequence is identified as the absence of emission in both the donor and the acceptor.

[0026]   In another embodiment, the method comprises hybridizing a nucleic acid primer comprising a donor fluorophore to a target nucleic acid at a primer binding site in the

15   target; exposing the hybridized primer to a first nucleotide comprising an acceptor fluorophore that, when incorporated into the primer, prevents further polymerization of the primer; detecting the presence or absence of fluorescent emission from each of the donor and the acceptor; identifying a nucleotide that has been incorporated into the primer via complementary base pairing with the target as the presence of fluorescent emission from the

20   acceptor; identifying a sequence placeholder as the absence of fluorescent emission from the donor and the acceptor; and repeating the exposing, detecting, and each of the identifying steps, thereby to compile a sequence of the target nucleic acid based upon the sequence of the incorporated nucleotides and the placeholders.

[0027]   The invention is useful in sequencing any form of polynucleotides, such as

25   double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA/RNA hybrids, RNAs with a recognition site for binding of the polymerizing agent, and RNA hairpins. The invention is particularly useful in high throughput sequencing of single molecule polynucleotides in which a plurality of target polynucleotides are attached to a solid support in a spatial arrangement such that each polynucleotides is individually optically

30   resolvable. According to the invention, each detected incorporated label represents a single polynucleotide.

[0028]   A detailed description of the certain embodiments of the invention is provided below. Other embodiments of the invention are apparent upon review of the detailed description that follows.

Brief Description of the Drawings

5   [0029]   The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

[0030]   Figure 1 shows asynchronous single molecule sequencing.

[0031]   Figure 2 are screenshots showing data from short cycle sequencing with long 10 homopolymer regions. Figure 2a shows full cycle sequencing used to analyze 10 target polynucleotides in a simulated synthesis of their complementary strands using cycle periods of 10 half-lives and repeating the wash cycles 12 times. Figure 2b shows a short cycle sequencing to analyze 10 target polynucleotides by simulating the synthesis of their complementary strands using short cycle periods of 0.8 half-life periods and repeating the 15 wash cycles 60 times.

[0032]   Figure 3 shows a short cycle embodiment for analyzing 200 target polynucleotides in a simulated synthesis of their complementary strands using short cycle periods of 0.8 half-life periods and repeating the wash cycles 60 times.

[0033]   Figure 4 shows a statistical analysis of incorporation, showing that 20 polymerizing agent may incorporate repeat labeled nucleotides less readily than the first labeled nucleotide.

[0034]   Figure 5 shows a simulation showing the effect of decreasing the activity rate of the polymerizing agent and lengthening half-lives on the cycle period.

[0035]   Figure 6 shows the number of cycles needed with cycle periods of various 25 half-lives taking into account stalling factors of two (squares), five (triangles) and 10 (crosses), in order to obtain over 25 incorporations in over 80% of target homopolymers, with at least a 97% chance of incorporating two or less nucleotides per cycle (or a smaller than 3% chance of incorporating more than 2 nucleotides per cycle).

[0036]   Figure 7 is a series of screenshots showing the effects of altering reaction 30 conditions on the incorporation of nucleotides in a single molecule sequencing by synthesis reaction.

Detailed Description

[0037]   The invention provides methods for high throughput single molecule sequencing. According to the invention, one or more parameters of a sequencing-by-synthesis reaction are preselected such that the incorporation of, preferably, a single nucleotide on a primed target template is optically detectable. In one embodiment, the preselected parameters

5   regulate the rate at which the nucleotides are incorporated, and the rate at which the incorporated nucleotides are detected. According to this embodiment, the nucleotides are individually detected either as they are incorporated or shortly thereafter, essentially in "real-time. In another embodiment, the preselected parameters permit the regulation of the number of nucleotides incorporated during a single extension cycle. In one aspect, the extension cycle

10   is stopped short at a predetermined point at which, on average, only 0, 1, 2, or 3 nucleotides have been incorporated into the primer, rather than permitting the reaction to run to near or full completion in each cycle.

[0038]   Short cycle methods according to the invention increase the resolution of individual nucleotides incorporated into the primer, but can decrease the yield of target

15   templates successfully incorporating a nucleotide in a single extension cycle. In traditional full cycle sequencing, nucleotides may be allowed to react in the presence of a polymerizing agent until at least one becomes incorporated into at least 99% of the complementary strands. This would produce a yield of $(0.99)^n \times 100\%$ for a complementary strand extended by n nucleotides. Obtaining incorporation in 99% of the complementary strands, however, requires

20   a period of several half-lives of the incorporation reaction, where one half-life is the time taken for at least one incorporation to occur in 50% of the complementary strands. Typically, the more strands that complete an incorporation during each cycle, the more n-mers obtained after n cycles.

[0039]   According to the invention, short cycle methods rely on a period of only a

25   limited number of half-lives of exposure to nucleotides, thus resulting in fewer target templates having incorporated a nucleotide in the short extension cycle. However, the short sequencing cycles provided by methods of the invention allow asynchronous analysis of polynucleotides. Thus, if an incorporation reactions fails to occur on a particular target polynucleotide, it can be completed in a later cycle without producing erroneous information,

30   or interfering with data from other target molecules being analyzed in parallel. As demonstrated in Figure 1, a cytosine ("C") incorporates into the extension product of one copy of a target polynucleotide, but fails to incorporate into the other copy. During subsequent cycles of incorporation, however, a C can be incorporated, without adversely affection

sequencing information. Thus, in asynchronous incorporation, an incorporation that failed to occur on a particular target in one-cycle can "catch up" in later cycles, permitting the use of shorter, even if more numerous, cycles.

[0040]   Because short cycle methods according the invention permit the detection of, for example, one, two or three individual nucleotides incorporated into a primed template, the invention overcomes the difficulty posed by homopolymer regions of a template sequence. While detection techniques may be able to quantify signal intensity from a smaller number of incorporated nucleotides of the same base-type, for example two or three incorporated nucleotides, longer runs of identical bases may not permit quantification due to increasing signal intensity. That is, it may become difficult to distinguish n bases from n+ 1 bases, where the fractional increase in signal intensity from the (n+ 1)'h base is small relative to the signal intensity from the already-incorporated n bases.

[0041]   In embodiments using short-cycles, it is possible to limit the number of nucleotides that become incorporated in a given cycle. For example, it can be determined by simulation that using a cycle period of about 0.8 half-lives can result in two or less incorporations in nine out of ten homopolymer complementary strands. (See Example 2b). In another simulation, a 0.8 half-life period was shown to allow no more than two incorporations in about 96.0% of 200 homopolymer complementary strands. As detection means can more readily quantify signal intensity from the smaller number of incorporated nucleotides rather than from larger numbers, the use of short-cycles addresses this issue. For example, imaging systems known in the art can reliably distinguish the difference in signal intensity between one versus two fluorescent labeling moieties on consecutively-incorporated nucleotides. Other imaging systems can reliably distinguish the difference in signal intensity between two versus three fluorescent labeling moieties on consecutively-incorporated nucleotides.

[0042]   In a further embodiment of the invention, an extension cycle comprising a labeled nucleotide is followed by an extension cycle using an unlabeled nucleotide of the same type so that the position in each of the target template in which a labeled nucleotide failed to incorporated becomes occupied by an unlabeled nucleotide. Methods in accordance with this embodiment provide for continued participation of specific template nucleic acids in which no incorporation of the labeled nucleotide occurred and reduced probability of missing nucleotides in the resulting compiled sequence.

[0043]   Further methods of the invention provide for identifying a placeholder in a nucleic acid sequence in the event that an accurate determination of a nucleotide at a particular

- 12 -

position is not possible. A placeholder is simply a position of unknown identity. Such a placeholder may be represented in a nucleic acid sequence with, for example, an "X," a traditional symbol for an unspecified nucleotide. Slotting a placeholder in a nucleic acid sequence avoids frameshift-type errors in sequence determination.

5   [0044]   Additional aspects of the invention are described in the following sections and illustrated by the Examples.

Target Nucleic Acids and Nucleotides

[0045]   The invention is useful in sequencing any form of polynucleotides, including double-stranded DNA, single-stranded DNA, single-stranded DNA hairpins, DNA/RNA

10   hybrids, RNAs with a recognition site for binding of the polymerizing agent, and RNA hairpins. Further, target polynucleotides may be a specific portion of a genome of a cell, such as an intron, regulatory region, allele, variant or mutation; the whole genome; or any portion therebetween. In other embodiments, the target polynucleotides may be mRNA, tRNA, rRNA, ribozymes, antisense RNA or RNAi. The target polynucleotide may be of any length,

15   such as at least 10 bases, at least 25 bases, at least 50 bases, at least 100 bases, at least 500 bases, at least 1000 bases, or at least 2500 bases. The invention is particularly useful in high throughput sequencing of single molecule polynucleotides in which a plurality of target polynucleotides are attached to a solid support in a spatial arrangement such that each polynucleotides is individually optically resolvable. According to the invention, each detected

20   incorporated label represents a single polynucleotide

[0046]   Nucleotides useful in the invention include both naturally-occurring and modified or non-naturally occurring nucleotides, and include nucleotide analogues. A nucleotide according to the invention may be, for example, a ribonucleotide, a deoxyribonucleotide, a modified ribonucleotide, a modified deoxyribonucleotide, a peptide

25   nucleotide, a modified peptide nucleotide or a modified phosphate-sugar backbone nucleotide. Many aspects of nucleotides useful in the methods of the invention are subject to manipulation provide and suitable mechanisms for controlling the reaction. In particular, the species or type of nucleotide (i.e., natural or synthetic dATP, dCTP, dTTP, dGTP or dUTP; a natural or non-natural nucleotide) will affect the rate or efficiency of the reaction and therefore require

30   consideration in preselecting parameters to produce the desire results.

[0047]   In addition, certain modifications to the nucleotides, including attaching a label, will affect the reaction. The size, polarity, hydrophobicity, hydrophilicity, charge, and other chemical attributes should be considered in determining parameters that will produce the

- 13 -

desired results in the reaction. Labeled nucleotides of the invention include any nucleotide that has been modified to include a label which is directly or indirectly detectable. Such labels include optically-detectable labels such fluorescent labels, including fluorescein, rhodamine, phosphor, polymethadine dye, fluorescent phosphoramidite, texas red, green fluorescent

5   protein, acridine, cyanine, cyanine 5 dye, cyanine 3 dye, 5-(2'-aminoethyl)-aminonaphthalene-
1-sulfonic acid (EDANS), BODIPY, ALEXA, or a derivative or modification of any of the foregoing. In one embodiment of the invention, fluorescence resonance energy transfer (FRET) technology is employed to produce a detectable, but quenchable, label. FRET may be used in the invention by, for example, modifying the primer to include a FRET donor moiety

10   and using nucleotides labeled with a FRET acceptor moiety.

[0048]   The fluorescently labeled nucleotides can be obtained commercially (e.g., from NEN DuPont, Amersham, and BDL). Alternatively, fluorescently labeled nucleotides can also be produced by various techniques, such as those described in Kambara et al., Bio/Techol. (1988) 6:816-821; Smith et al., Nucl. Acid Res. (1985) 13: 2399-2412, and Smith

15   et al.., Nature (1986) 321: 674-79.

[0049]   The fluorescent dye is preferably linked to the deoxyribose by a linker arm which is easily cleaved by chemical or enzymatic means. The length of the linker between the dye and the nucleotide can impact the incorporation rate and efficiency (see Zhu et al., Cytometry (1997) 28, 206). There are numerous linkers and methods for attaching labels to

20   nucleotides, as shown in Oligonucleotides and Analogues: A Practical Approach (1991) (IRL Press, Oxford); Zuckerman et al., Polynucleotides Research (1987) 15: 5305-21; Sharma et al., Polynucleotides Research, (1991) 19: 3019; Giusti et al., PCR Methods and Applications (1993) 2: 223-227; Fung et al., U.S. Patent No. 4,757,141; Stabinsky, U.S. Patent No. 4, 739,044; Agrawal et al., Tetrahedron Letters, (1990) 31: 1543-46; Sproat et al.,

25   Polynucleotides Research (1987) 15: 4837; and Nelson et al., Polynucleotides Research, (1989) 17: 7187-94.

[0050]   While the invention is exemplified herein with fluorescent labels, the invention is not so limited and can be practiced using nucleotides labeled with any form of detectable label, including radioactive labels, chemoluminescent labels, luminescent labels,

30   phosphorescent labels, fluorescence polarization labels, and charge labels.

Reaction Parameters

[0051]   Any parameter that permits the regulation of the number of labeled nucleotides added to the primer, or the rate at which the nucleotides are incorporated and

detected can be controlled or exploited in the practice of the invention. Such parameters include, for example, the presence or absence of a label on a nucleotide, the type of label and manner of label attachment; the linker identity and length used to attach the label; the type of nucleotide (including, for example, whether such nucleotide is a dATP, dCTP, dTTP, dGTP or dUTP; a natural or non-natural nucleotide, a nucleotide analogue, or a modified nucleotide); the local sequence immediately 3' to the addition position; whether the base is the first, second, third, etc. base added; the type of polymerase used; the particular batch characteristics of the polymerase; the processivity of the polymerase; the incorporation rate of the polymerase, and use of polymerase cofactors.

[0052]   In addition, a variety of the conditions of the reaction provide useful mechanisms for controlling either the number of nucleotides incorporated in a single extension reaction or the rates of nucleotide incorporation and detection. Such conditions include the "half-life" of the extension cycle (where one half-life is the time taken for at least one incorporation to occur in 50% of the complementary strands); the number of wash cycles (i.e., the number of times a nucleotide is introduced to the reaction then washed out); the number of target nucleic acids in the reaction; and the temperature of the reaction and the reagents used in the reaction.

Half-Lives and Wash Cycles

[0053]   Based on the methods disclosed herein, those of skill in the art will be able to determine the period of half-lives required to limit the number incorporations per cycle for a given number of target polynucleotides. (See Examples 2 and 3, Figures 2 and 3). Statistical simulations can also provide the number of repeated cycles needed to obtain a given number of incorporations, for example, to sequence a 25 base pair sequence. (See Examples 2 and 3, Figures 2 and 3). Referring to the simulations above, for example, it can be determined that 60 cycles, each 0.8 half-lives long, would be required for at least 25 incorporations in each of ten complementary strands (Example 2b, Figure 2b). With 200 complementary strands, 60 cycles each 0.8 half-lives long produce at least 20 incorporations in each strand (Example 3, Figure 3). Following the methodologies outlined herein, such as the simulated working examples detailed below, those of skill in the art will be able to make similar determinations for other numbers of targets of varying lengths, and use appropriate cycle periods and numbers of cycles to analyze homopolymer without using blocking moieties or reversible chain termination.

[0054]   The cycle period may also be chosen to permit a certain chance of

incorporation of a given number of nucleotides in a complementary strand, and the cycle may be repeated a number of times to analyze the sequence of various numbers of target polynucleotides of varying length.

[0055]   In some embodiments, nucleotide half-lives for the incorporation reaction are affected by the fact that polymerizing agent may incorporate labeled nucleotides less readily than unlabeled nucleotides. Figure 4 illustrates the statistics of incorporation for a certain embodiment using a Klenow exo-minus polymerizing agent and Cy3- or Cy5- labeled nucleotides. The results show that polymerase may incorporate subsequent labeled nucleotides less readily than a prior labeled nucleotide. The graph of Figure 4 indicates, for example, that it may take five to ten times longer, resulting in a "stalling" of the incorporation reaction. In other embodiments, the stalling may vary with the use of other labeled nucleotides, other polymerizing agents and various reaction conditions.

[0056]   Polymerase stalling is a useful mechanism for controlling incorporation rates in single molecule reactions. As is shown in the Examples below, polymerase stalling is useful to limit incorporation of nucleotides into any given strand in a fairly precise manner. According to the invention, polymerase stalling is useful to limit incorporation to 1 nucleotide per strand per cycle, on average. Given a priori knowledge of the statistics of incorporation, single molecule reactions are controlled to provide a statistical likelihood that 1, sometimes 2, but rarely 3 nucleotides are incorporated in a strand in any given cycle.

[0057]   For example, the rate at which polymerase incorporates labeled nucleotides into a complementary strand may be slowed by a factor of about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, or about 15 times compared to that observed with unlabeled nucleotides or compared to that observed for a prior incorporated labeled nucleotide.

[0058]   Moreover, this inhibition or delaying and longer half-lives can be taken into account when determining appropriate cycle periods and numbers of cycles to analyze homopolymer targets of a given length. Figures 3 and 4, for example, illustrate the results of simulations in which various factors affecting incorporation rates are taken into account. The graph of Figure 4, for example, shows the number of cycles needed with cycle periods of various half-lives, taking into account stalling factors of two (squares), five (triangles), and 10 (crosses), in order to obtain 25 incorporations in over 80% of target strands, with at least a 97% chance of incorporating two or fewer nucleotides per cycle (or a smaller than 3% chance of incorporating three or more nucleotides per cycle). As the graph shows, stalling allows

longer half-lives, which, in turn, permits the use of fewer cycles to obtain a "full" sequence with a defined error rate. As Figure 5 illustrates, if the use of labeled nucleotides slows down the polymerizing agent by a factor of 5, a cycle period of 2.4 half-lives produces over 80% 25-mers in 30 cycles. Based on the teachings of the invention, one of ordinary skill in the art

5   can determine the cycle period required to limit the number incorporations per cycle for a given number of target polynucleotides of a given length.

[0059]   Applying methods disclosed herein, the cycle period may be selected to permit about a 70%, about a 75%, about an 80%, about an 85%, about a 90%, about a 95%, about a 96%, about a 97%, about a 98%, and about a 99% chance of incorporation of two or less

10   nucleotides into the complementary strand. Other cycle periods that may be used in embodiments of the invention include, for example, no more than about 5 half-lives, no more than about 4 half-lives, no more than about 3 half-lives, no more than about 2 half-lives, no more than about 1 half-lives, no more than about 0.9 half-lives, no more than about 0.8 half-lives, no more than about 0.7 half-lives, no more than about 0.6 half-lives, no more than about

15   0.5 half-lives, no more than about 0.4 half-lives, no more than about 0.3 half-lives, and no more than about 0.2 half-lives of the incorporation reactions.

[0060]   In addition to the Examples provided below, various cycle periods and number of times the cycles are repeated may be used with various numbers of targets in certain embodiments of the invention. These include, for example, using about 200 target

20   polynucleotides, a period of no more than about 0.6 half-lives and repeating at least about 50 times; using about 200 target polynucleotides, a period of no more than about 0.6 half-lives and repeating at least about 60 times; using about 200 target polynucleotides, a period of no more than about 0.6 half-lives and repeating at least about 70 times; using about 200 target polynucleotides, a period of no more than about 0.8 half-lives and repeating at least about 50

25   times; using about 200 target polynucleotides, a period of no more than about 0.8 half-lives and repeating at least about 60 times; using about 200 target polynucleotides, a period of no more than about 0.8 half-lives and repeating at least about 70 times; using about 200 target polynucleotides, a period of no more than about 1 half-life and repeating at least about 50 times; using about 200 target polynucleotides, a period of no more than about 1 half-life and

30   repeating at least about 60 times; and using about 200 target polynucleotides, a period of no more than about 1 half-life and repeating at least about 70 times. In any of these embodiments, signal from incorporated nucleotides may be reduced after each or a number of cycles.

[0061]   The number of times the cycles need to be repeated is also determined based on methods described herein. In general, the number of cycles increases with the length of the sequence to be analyzed and the duration of the half life of nucleotide exposure decreases as the length of sequence to be analyzed becomes longer. Also in general, half lives of

5   nucleotide exposure increase and cycle numbers decrease with greater inhibitory or delaying effects on nucleotide incorporation

[0062]   Taking into account various stalling factors, examples of cycle periods and number repeat cycles that may be used in certain embodiments further include a cycle period of no more than about 0.5 half-lives with a stalling factor of about 2, repeated at least about 90

10   times; a cycle period of no more than about 0.75 half-lives, with a stalling factor of about 2, repeated at least about 75 times; a cycle period of no more than about 1 half-lives, with a stalling factor of about 2, repeated at least about 50 times; a cycle period of no more than about 1.5 half-lives with a stalling factor of about 2 or about 5, repeated at least about 45 times; a cycle period of no more than about 1.75 half-lives, with a stalling factor of about 5,

15   repeated at least about 35 times; a cycle period of no more than about 2 half-lives, with a stalling factor of about 5 or about 10, repeated at least about 35 times; a cycle period of no more than about 2.25 half-lives, with a stalling factor of about 5 or about 10, repeated at least about 30 or at least about 35 times, and a cycle period of about 2.4 half-lives, with a stalling factor of about 5, repeated at least about 30 times.

20   *Polymerases and Polymerase Cofactors*

[0063]   Polymerizing agents useful in the invention include DNA polymerases (such as Taq polymerase, T7 mutant DNA polymerase, Klenow and Sequenase, 9°N or a variant thereof), RNA polymerases, thermostable polymerases, thermodegradable polymerases, and reverse transcriptases. See *e.g.,* Doublie et al., Nature (1998) 391:251-58; Ollis et al. Nature

25   (1985) 313: 762-66; Beese et al., Science (1993) 260: 352-55; Korolev et al., Proc. Natl. Acad. Sci. USA (1995) 92: 9264-68; Keifer et al., Structure (1997) 5:95-108; and Kim et al., Nature (1995) 376:612-16.

[0064]   Cofactors of the invention function to inhibit the polymerizing agent, thereby slowing or stopping synthesis activity, permitting the detection of an incorporated labeled

30   nucleotide. Cofactors of the invention include any chemical agent or reaction condition that results in the inhibition of the polymerizing agent. Such inhibition may be in whole or in part and may be permanent, temporary or reversible. For example, a cofactor may be a label, an antibody, an aptamer, an organic or inorganic small molecule, or a polyanion, or it may

- 18 -

comprise a chemical modification to a nucleotide (*i.e.*, a nucleotide analogue may comprise a cofactor). A cofactor can be in solution, or it may be attached, either directly or through a linker to a nucleotide, primer, template or polymerase.

[0065]   Examples of useful cofactor agents include, among others, light sensitive groups such as 6-nitoveratryloxycarbonyl (NVOC), 2-nitobenzyloxycarbonyl (NBOC), α, α-dimethyl-dimethoxybenzyloxycarbonyl (DDZ), 5-bromo-7-nitroindolinyl, o-hyrdoxy-2-methyl cinnamoyl, 2-oxymethylene anthraquinone, and t-butyl oxycarbonyl (TBOC). Oligonucleotide Synthesis: A Practical Approach (IRL Press, Oxford). Useful polyanions are described in U.S. Patent No. 6,667,165 (the disclosure of which is incorporated by reference herein); and useful aptamers are described in U.S. Patent Nos. 6,020,130 and 6,183,967 (the disclosures of which are incorporated by reference herein). See U.S. Patent No. 5,338,671 for useful antibodies. Nucleotides possessing various labels and cofactors can be readily synthesized. Labeling moieties are attached at appropriate sites on the nucleotide using chemistry and conditions as described in Gait (1984).

[0066]   Further, the cofactor may also be the detectable label. Labels useful as combined labels/cofactors include larger or bulky dyes. For example, the detectable label may comprise a dye having a bulky chemical structure that, once the nucleotide is incorporated into the extending primer, causes a steric hindrance of the polymerizing agent, blocking the polymerizing agent from any further synthesis. Examples of labels that may be useful for this purpose are described in the Example, as well as in Zhu et al., Polynucleotides Res. (1994) 22: 3418-22. For example, fluorophore labels that may be used to stall the polymerase include Cy3, Cy5, Cy7, ALEXA647, ALEXA 488, BODIPY 576/589, BODIPY 650/665, BODIPY TR, Nile Blue, Sulfo-IRD700, NN382, R6G, Rho123, tetramethylrhodamine and Rhodamine X. In one embodiment, the labels are as bulky as Cy5, with molecular weights at least about 1.5 kDa. In another embodiment, the labels are bulkier than Cy5, having molecular weights of at least about 1.6 kDa, at least about 1.7 kDa, at least about 1.8 kDa, at least about 1.9 kDa, at least about 2.0 kDa at least bout 2.5 kDa, or at least about 3.0 kDa.

[0067]   Further examples of such larger dyes include the following, with corresponding formula weights (in g/mol) in parentheses: Cy5 (534.6); Pyrene (535.6); 6-Carboxyfluorescein (FAM) (537.5); 6-Carboxyfluorescein-DMT (FAM-X (537.5); 5(6) Carboxyfluorescein (FAM) (537.5); 5-Fluorescein (FITC) (537.6); Cy3B (543.0); WellRED D4-PA (544.8); BODIPY 630/650 (545.5); 3' 6-Carboxyfluorescein (FAM) (569.5); Cy3.5 (576.7); Cascade Blue (580.0); ALEXA Fluor 430 (586.8); Lucifer Yellow (605.5); ALEXA

- 19 -

Fluor 532 (608.8); WellRED D2-PA (611.0); Cy5.5 (634.8); DY-630 (634.8); DY-555
(636.2); WellRED D3-PA (645.0); Rhodamine Red-X (654.0); DY-730 (660.9); DY-782
(660.9); DY-550 (667.8); DY-610 (667.8); DY-700 (668.9); 6-Tetrachlorofluorescein (TET)
(675.2) ALEXA Fluor 568 (676.8); DY-650 (686.9); 5(6)- Carboxyeosin (689.0); Texas Red-

5   X (702.0); ALEXA Fluor 594 (704.9); DY-675 (706.9); DY-750 (713.0); DY-681 (736.9);
Hexachlorofluorescein (HEX) (744.1); DY-633 (751.9); LightCycler Red 705 (753.0);
LightCycler Red 640 (758.0); DY-636 (760.9); DY-701 (770.9); FAR-Fuchsia (5'-Amidite)
(776.0); FAR-Fuchsia (SE) (776.0); DY-676 (808.0); Erythrosin (814); FAR-Blue (5'-
Amidite) (824.0); FAR-Blue (SE) (824.0); Oyster 556 (850.0); Oyster 656 (900.0); FAR-

10   Green Two (SE) (960.0); ALEXA Fluor 546 (964.4); FAR-Green One (SE), (976.0); ALEXA
Fluor 660 (985.0); Oyster 645 (1000.0); ALEXA Fluor 680 (1035.0); ALEXA Fluor 633
(1085.0); ALEXA Fluor 555 (1135.0); ALEXA Fluor 647 (1185.0); ALEXA Fluor 750
(1185.0); ALEXA Fluor 700 (1285.0). These reagents are commercially available from
SYNTHEGEN, LLC (Houston, Tex.).

15   [0068]   There is extensive guidance in the literature for derivatizing fluorophore and
quencher molecules for covalent attachment via common reactive groups that can be added to
a nucleotide (see Haugland, Handbook of Fluorescent Probes and Research Chemicals (1992).
There are also many linking moieties and methods for attaching fluorophore moieties to
nucleotides, as described in Oligonucleotides and Analogues, *supra*; Guisti et al., supra;

20   Agrawal et al, Tetrahedron Letters (1990) 31: 1543-46; and Sproat et al., Polynucleotide
Research (1987) 15: 4837.

[0069]   In one embodiment, the method further comprises inactivating the cofactor,
thereby reversing its effect on the polymerizing agent. Modes of inactivation depend on the
cofactor. For example, where the cofactor is attached to the nucleotide, inactivation can

25   typically be achieved by chemical, enzymatic, photochemical or radiation cleavage of the
cofactor from the nucleotide. Cleavage of the cofactor can be achieved if a detachable
connection between the nucleotide and the cofactor is used. For example, the use of disulfide
bonds enables one to disconnect the dye by applying a reducing agent like dithiothreitol
(DTT). In a further alternative, where the cofactor is a fluorescent label, it is possible to

30   neutralize the label by bleaching it with radiation.

[0070]   In the event that temperature-sensitive cofactors are utilized, inactivation may
comprise adjusting the reaction temperature. For example, an antibody that binds to
thermostable polymerase at lower temperatures and blocks activity, but is denatured at higher

- 20 -

temperatures, thus rendering the polymerase active; or single-stranded aptamers that bind to thermophilic polymerase at lower temperatures but are released at higher temperatures, may be inactivated by increasing the reaction temperature such the cofactor is released but polymerase activity is permitted.

5   [0071]   In one embodiment, transient inhibition of the polymerase and the time of exposure to the labeled nucleotide are coordinated such that it is statistically likely that at least one of the labeled nucleotide is incorporated in the primer, but statistically unlikely that more than two of the labeled nucleotide are incorporated. In another embodiment, the reaction is controlled by inhibiting the polymerase activity such that it is statistically unlikely that more

10   than, for example, one or two nucleotides are incorporated into the same primer strand in the cycle.

Temperature and Reagents

[0072]   Other reaction conditions that are useful in the methods of the invention include reaction temperature and reagents. For example, a temperature above or below the

15   temperature required for optimal activity of the polymerizing agent, such as a temperature of about 20-70°, would be expected to result in a modulation of the polymerization rate, C. This form of inhibition is typically reversible with correction of the reaction temperature, provided that the delta in temperature was insufficient to cause a permanent damage to the polymerase.

[0073]   In another embodiment, buffer reagents useful in the methods of the invention

20   include a detergent or surfactant, such as Triton®-X 100, or salt and/or ion concentrations that facilitate or inhibit nucleotide incorporation.

Predetermined Points For Stopping a Cycle

[0074]   The predetermined point at which a short cycle is stopped is defined, for example, by the occurrence of an event (such as the incorporation of a nucleotide comprising

25   a blocking moiety that prevents further extension of the primer), the lapse of a certain amount of time (such as a specific number of half-lives), or the achievement of a statistically-significant datapoint (such as a period at which a statistically significant probability of two or less nucleotides have been incorporated). In one embodiment, the predetermined period of time is coordinated with an amount of polymerization inhibition such that, on average, a

30   certain number of labeled nucleotides are added to the primer. In another embodiment, the number of incorporated labeled nucleotides is, on average, 0, 1 or 2, but almost never more than 3. The time period of exposure is defined in terms of statistical significance. For example, the time period may be that which is statistically insufficient for incorporation of

more nucleotides than are resolvable by a detection system used to detect incorporation of the nucleotide into the primer. In another example, the time period that is statistically insufficient for incorporation of a greater number of nucleotides that are individually optically resolvable during a predetermined detection period (i.e., a period of time during which the incorporated

5   nucleotides are detected).

[0075]   The reaction may be stopped by washing or flushing out the nucleotides that remain unincorporated and/or washing or flushing out polymerization agent. Further, many aspects of the repeated cycles may be automated, for example, using microfluidics for washing nucleotides to sites of anchored target polynucleotides, and washing out

10   unincorporated nucleotides to halt each cycle.

[0076]   The following exemplifications of the invention are useful in understanding certain aspects of the invention but are not intended to limit the scope of the invention in any way.

15   Example 1

[0077]   Primers are synthesized from nucleoside triphosphates by known automated oligonucleotide synthetic techniques, *e.g.*, via standard phosphoramidite technology utilizing a nucleic acid synthesizer, such as the ABI3700 (Applied Biosystems, Foster City, CA). The oligonucleotides are prepared as duplexes with a complementary strand, however, only the 5'

20   terminus of the oligonucleotide proper (and not its complement) is biotinylated.

Ligation of Oligonucleotides and Target polynucleotides

[0078]   Double stranded target nucleic acids are blunt-end ligated to the oligonucleotides in solution using, for example, T4 ligase. The single strand having a 5' biotinylated terminus of the oligonucleotide duplex permits the blunt-end ligation on only on

25   end of the duplex. In a preferred embodiment, the solution-phase reaction is performed in the presence of an excess amount of oligonucleotide to prohibit the formation of concantamers and circular ligation products of the target nucleic acids. Upon ligation, a plurality of chimeric polynucleotide duplexes result. Chimeric polynucleotides are separated from unbound oligonucleotides based upon size and reduced to single strands by subjecting them to

30   a temperature that destabilizes the hydrogen bonds.

Preparation of Solid Support

[0079]   A solid support comprising reaction chambers having a fused silica surface is sonicated in 2% MICRO-90 soap (Cole-Parmer, Vernon Hills, IL) for 20 minutes and then

cleaned by immersion in boiling RCA solution (6:4:1 high-purity H_2O/30% NH_4OH/30% H_2O_2) for 1 hour. It is then immersed alternately in polyallylamine (positively charged) and polyacrylic acid (negatively charged; both from Aldrich) at 2 mg/ml and pH 8 for 10 minutes each and washed intensively with distilled water in between. The slides are incubated with 5 mM biotin-amine reagent (Biotin-EZ-Link, Pierce) for 10 minutes in the presence of 1-[3-(dimethylamino)propyl]-3-ethylcarbodiimide hydrochloride (EDC, Sigma) in MES buffer, followed by incubation with Streptavidin Plus (Prozyme, San Leandro, CA) at 0.1 mg/ml for 15 minutes in Tris buffer. The biotinylated single-stranded chimeric polynucleotides are deposited via ink-jet printing onto the streptavidin-coated chamber surface at 10 pM for 10 minutes in Tris buffer that contain 100 mM $MgCl_2$.

Equipment

[0080]   The experiments are performed on an upright microscope (BH-2, Olympus, Melville, NY) equipped with total internal reflection (TIR) illumination, such as the BH-2 microscope from Olympus (Melville, NY). Two laser beams, 635 (Coherent, Santa Clara, CA) and 532 nm (Brimrose, Baltimore), with nominal powers of 8 and 10 mW, respectively, are circularly polarized by quarter-wave plates and undergo TIR in a dove prism (Edmund Scientific, Barrington, NJ). The prism is optically coupled to the fused silica bottom (Esco, Oak Ridge, NJ) of the reaction chambers so that evanescent waves illuminated up to 150 nm above the surface of the fused silica. An objective (DPlanApo, 100 UV 1.3oil, Olympus) collects the fluorescence signal through the top plastic cover of the chamber, which is deflected by the objective to ≈40 μm from the silica surface. An image splitter (Optical Insights, Santa Fe, NM) directs the light through two bandpass filters (630dcxr, HQ585/80, HQ690/60; Chroma Technology, Brattleboro, VT) to an intensified charge-coupled device (I-PentaMAX; Roper Scientific, Trenton, NJ), which records adjacent images of a 120- × 60-μm section of the surface in two colors.

Experimental Protocols

FRET-Based Method Using Nucleotide-Based Donor Fluorophore

[0081]   In a first experiment, universal primer is hybridized to a primer attachment site present in support-bound chimeric polynucleotides. Next, a series of incorporation reactions are conducted in which a first nucleotide comprising a cyanine-3 donor fluorophore is incorporated into the primer as the first extended nucleotide. If all the chimeric sequences are the same, then a minimum of one labeled nucleotide must be added as the initial FRET donor because the template nucleotide immediately 3' of the primer is the same on all chimeric

polynucleotides. If different chimeric polynucleotides are used (i.e., the polynucleotide portion added to the bound oligonucleotides is different at least one location), then all four labeled dNTPs initially are cycled. The result is the addition of at least one donor fluorophore to each chimeric strand.

5   [0082]   The number of initial incorporations containing the donor fluorophore is limited by either limiting the reaction time (i.e., the time of exposure to donor-labeled nucleotides), by polymerase stalling, or both in combination. The inventors have shown that base-addition reactions are regulated by controlling reaction conditions. For example, incorporations can be limited to 1 or 2 at a time by causing polymerase to stall after the

10   addition of a first base. One way in which this is accomplished is by attaching a dye to the first added base that either chemically or sterically interferes with the efficiency of incorporation of a second base. A computer model was constructed using Visual Basic (v. 6.0, Microsoft Corp.) that replicates the stochastic addition of bases in template-dependent nucleic acid synthesis. The model utilizes several variables that are thought to be the most

15   significant factors affecting the rate of base addition. The number of half-lives until dNTPs are flushed is a measure of the amount of time that a template-dependent system is exposed to dNTPs in solution. The more rapidly dNTPs are removed from the template, the lower will be the incorporation rate. The number of wash cycles does not affect incorporation in any given cycle, but affects the number bases ultimately added to the extending primer. The number of

20   strands to be analyzed is a variable of significance when there is not an excess of dNTPs in the reaction. Finally, the inhibition rate is an approximation of the extent of base addition inhibition, usually due to polymerase stalling. The homopolymer count within any strand can be ignored for purposes of this application. Figure 2 is a screenshot showing the inputs used in the model.

25   [0083]   The model demonstrates that, by controlling reaction conditions, one can precisely control the number of bases that are added to an extending primer in any given cycle of incorporation. For example, as shown in Figure 7, at a constant rate of inhibition of second base incorporation (i.e., the inhibitory effect of incorporation of a second base given the presence of a first base), the amount of time that dNTPs are exposed to template in the

30   presence of polymerase determines the number of bases that are statistically likely to be incorporated in any given cycle (a cycle being defined as one round of exposure of template to dNTPs and washing of unbound dNTP from the reaction mixture). As shown in Figure 7a, when time of exposure to dNTPs is limited, the statistical likelihood of incorporation of more

than two bases is essentially zero, and the likelihood of incorporation of two bases in a row in
the same cycle is very low. If the time of exposure is increased, the likelihood of
incorporation of multiple bases in any given cycle is much higher. Thus, the model reflects
biological reality. At a constant rate of polymerase inhibition (assuming that complete stalling

5   is avoided), the time of exposure of a template to dNTPs for incorporation is a significant
factor in determining the number of bases that will be incorporated in succession in any cycle.
Similarly, if time of exposure is held constant, the amount of polymerase stalling will have a
predominant effect on the number of successive bases that are incorporated in any given cycle
(See, Figure 7b). Thus, it is possible at any point in the sequencing process to add or renew

10   donor fluorophore by simply limiting the statistical likelihood of incorporation of more than
one base in a cycle in which the donor fluorophore is added.

[0084]   Upon introduction of a donor fluorophore into the extending primer sequence,
further nucleotides comprising acceptor fluorophores (here, cyanine-5) are added in a
template-dependent manner. It is known that the Foster radius of Cy-3/Cy5 fluorophore pairs

15   is about 5 nm (or about 15 nucleotides, on average). Thus, donor must be refreshed about
every 15 bases. This is accomplished under the parameters outlined above. In general, each
cycle preferably is regulated to allow incorporation of 1 or 2, but never 3 bases. So,
refreshing the donor means simply the addition of all four possible nucleotides in a mixed-
sequence population using the donor fluorophore instead of the acceptor fluorophore every

20   approximately 15 bases (or cycles). Figure 2 shows schematically the process of FRET-based,
template-dependent nucleotide addition as described in this example.

[0085]   The methods described above are alternatively conducted with the FRET
donor attached to the polymerase molecule. In that embodiment, donor follows the extending
primer as new nucleotides bearing acceptor fluorophores are added. Thus, there typically is

25   no requirement to refresh the donor. In another embodiment, the same methods are carried
out using a nucleotide binding protein (e.g., DNA binding protein) as the carrier of a donor
fluorophore. In that embodiment, the DNA binding protein is spaced at intervals (e.g., about 5
nm or less) to allow FRET. Thus, there are many alternatives for using FRET to conduct
single molecule sequencing using the devices and methods taught in the application.

30   However, it is not required that FRET be used as the detection method. Rather, because of the
intensities of the FRET signal with respect to background, FRET is an alternative for use
when background radiation is relatively high.

   Non-FRET Based Methods

[0086]   Methods for detecting single molecule incorporation without FRET are also conducted. In this embodiment, incorporated nucleotides are detected by virtue of their optical emissions after sample washing. Primers are hybridized to the primer attachment site of bound chimeric polynucleotides Reactions are conducted in a solution comprising Klenow fragment Exo-minus polymerase (New England Biolabs) at 10 nM (100 units/ml) and a labeled nucleotide triphosphate in EcoPol reaction buffer (New England Biolabs). Sequencing reactions takes place in a stepwise fashion. First, 0.2 μM dUTP-Cy3 and polymerase are introduced to support-bound chimeric polynucleotides, incubated for 6 to 15 minutes, and washed out. Images of the surface are then analyzed for primer-incorporated U-Cy5. Typically, eight exposures of 0.5 seconds each are taken in each field of view in order to compensate for possible intermittency (e.g., blinking) in fluorophore emission. Software is employed to analyze the locations and intensities of fluorescence objects in the intensified charge-coupled device pictures. Fluorescent images acquired in the WinView32 interface (Roper Scientific, Princeton, NJ) are analyzed using ImagePro Plus software (Media Cybernetics, Silver Springs, Md). Essentially, the software is programmed to perform spot-finding in a predefined image field using user-defined size and intensity filters. The program then assigns grid coordinates to each identified spot, and normalizes the intensity of spot fluorescence with respect to background across multiple image frames. From those data, specific incorporated nucleotides are identified. Generally, the type of image analysis software employed to analyze fluorescent images is immaterial as long as it is capable of being programmed to discriminate a desired signal over background. The programming of commercial software packages for specific image analysis tasks is known to those of ordinary skill in the art. If U-Cy5 is not incorporated, the substrate is washed, and the process is repeated with dGTP-Cy5, dATP-Cy5, and dCTP-Cy5 until incorporation is observed. The label attached to any incorporated nucleotide is neutralized, and the process is repeated. To reduce bleaching of the fluorescence dyes, an oxygen scavenging system can be used during all green illumination periods, with the exception of the bleaching of the primer tag.

[0087]   In order to determine a template sequence, the above protocol is performed sequentially in the presence of a single species of labeled dATP, dGTP, dCTP or dUTP. By so doing, a first sequence can be compiled that is based upon the sequential incorporation of the nucleotides into the extended primer. The first compiled sequence is representative of the complement of the template. As such, the sequence of the template can be easily determined by compiling a second sequence that is complementary to the first sequence. Because the

sequence of the oligonucleotide is known, those nucleotides can be excluded from the second sequence to produce a resultant sequence that is representative of the target template.

Example 2

[0088]   Figure 2 illustrates the advantage of short-cycle sequencing with respect to avoiding long homopolymer reads. Figure 2a illustrates a simulated analysis of 10 target polynucleotides using non-short-cycle sequencing (Example 2a), whereas Figure 2b illustrates a simulated analysis of the same number of target polynucleotides using short-cycle sequencing (Example 2 b).

[0089]   The simulations were performed as follows: an Excel spreadsheet was opened and "Customize..." selected from the "Tools" menu of the Excel toolbar. The "Commands" tab was selected and, after scrolling down, "Macros" was clicked. The "smiley face" that appeared in the right panel was dragged to the toolbars on top of the spreadsheet. The "Customize" box was closed and the "smiley face" clicked once. From the list of subroutines that appeared, "ThisWorkbook.Main_Line." was selected. The program was run by clicking again on the "smiley face." A copy of the source code for the Excel simulation is provided below.

[0090]   Input values were then entered into the tabbed sheet called "In Out." There were three input values:

[0091]   The first input value corresponded to the period of time allowed for incorporation reactions of provided nucleotides into the growing complementary strands of the polynucleotides to be analyzed. This period was conveniently measured in half-lives of the incorporation reaction itself. Each cycle of incorporation was simulatedly stalled after a period of time, representing, for example, the time when unincorporated nucleotides would be flushed out or the incorporation reactions otherwise stalled.

[0092]   The second input value corresponds to the number of times each cycle of incorporation was repeated. That is, the number of times the steps of providing nucleotides, allowing incorporation reactions into the complementary strands in the presence of polymerizing agent, and then stopping the incorporations are repeated. The nucleotides were simulatedly provided as a wash of each of dATPs, dGTPs, dTTPs, and dCTPs. The program then recorded which nucleotides were incorporated, corresponding to a detection step of detecting incorporation.

[0093]   The third input value corresponds to number of strands of target polynucleotides to by analyzed in the simulation. The program allowed up to 1100 target polynucleotide molecules to be analyzed in a given simulation.

[0094]   After the program was started, as described above, the program first generated

- 28 -

the inputted number of strands composed of random sequences. The program then simulated hybridization and polymerization of the correct base of each incorporation reaction, based on the generated sequence of the target polynucleotide templates. The program continued these simulated reactions for the allowed amount of simulated time, determined by the inputted

5   number of half-lives. Statistics of the simulation were then computed and reported, including the longest strand, the shortest strand, and the average length of all strands, as well as the fraction of strands extended by at least 25 nucleotide incorporations, as discussed in more detail below.

[0095]   In the first part of this simulation, Example 2 a, the input values used were a

10   cycle period of 10 half-lives, 12 repeats of the cycle, and 10 target polynucleotide strands.

[0096]   Figure 2a illustrates the results obtained. Homopolymers stretches which occurred in the same simulated complementary strand are highlighted in magenta wherever 2 nucleotides of the same base type were incorporated in a row, and in cyan wherever more than two nucleotides of the same base type were incorporated in a row.

15   [0097]   Figure 2a illustrates that the output values included the longest extended complementary strand obtained during the simulation (Longest extension in the ensemble of molecules); the shorted extended complementary strand obtained during the simulation (Shortest extension in the ensemble of molecules); and the average extension. These numbers represent the greatest number of incorporations into any of the 10 simulatedly

20   growing complementary strands, the smallest number of incorporations for any of the 10, and the average number of incorporations for the 10. Figure 2a indicates that the values obtained for Example 2a were 37 incorporations in the longest extension, 25 in the shortest, and 30.00 as the average number of incorporations.

[0098]   The output values also provided information on the number of incorporations

25   that occurred in each of growing complementary strands during each cycle period of the simulation. For example, Figure 2a indicates that for the input values of Example 2a, the percentage of growing stands extended by two or more nucleotides in a homopolymer stretch was 100.0%; and the percentage of growing strands extended by three or more nucleotides in a homopolymer stretch was 60.0%. That is, using a cycle period of 10 half-lives resulted in

30   only 40% of the complementary strands being extended by two or less nucleotides in a homopolymer stretch per cycle of incorporation.

[0099]   Further, output values also indicated the total number of incorporations for each of the growing strands for the total number of repeated cycles. This represents the

length of the sequence of target polynucleotide analyzed. Figure 2a illustrates that in Example 2 a, 100.0% of the 10 target polynucleotides of the simulation were extended by at least 25 incorporated nucleotides. This illustrates that using a cycle period of 10 half-lives, and repeating the cycles 12 times, allowed analysis of a 25 base sequence of 10 target polynucleotides.

[0100]   In the second part of this simulation, Example 2b, the input values used were a cycle period of 0.8 half-lives, 60 repeats of the cycle, and 10 target polynucleotide strands.

[0101]   Figure 2b illustrates the results obtained. Homopolymers stretches which occurred in the same simulated complementary strand are highlighted in magenta wherever 2 nucleotides of the same base type were incorporated in a row, and in cyan wherever more than two nucleotides of the same base type were incorporated in a row.

[0102]   Figure 2b illustrates that the output values included the longest extended complementary strand obtained during the simulation (longest extension in the ensemble of molecules); the shortest extended complementary strand obtained during the simulation (shortest extension in the ensemble of molecules); and the average extension. These numbers represent the greatest number of incorporations into any of the 10 simulatedly growing complementary strands, the smallest number of incorporations for any of the 10, and the average number of incorporations for the 10. Figure 2b indicates that the values obtained for Example 2b were 37 incorporations in the longest extension, 26 in the shortest, and 32.00 as the average number of incorporations.

[0103]   The output values also provided information on the number of incorporations that occurred in each of growing complementary strands during each cycle period of the simulation. For example, Figure 2b indicates that for the input values of Example 2b, the percentage of growing stands extended by two or more nucleotides in a homopolymer stretch was 80.0%; and the percentage of growing strands extended by three or more nucleotides in a homopolymer stretch was 10.0%. That is, using a cycle period of 0.8 half-lives resulted in 90% of the complementary strands being extended by two or less nucleotides per cycle of incorporation.

[0104]   Output values also indicated the total number of incorporations for each of the growing strands for the total number of repeated cycles. As in Example 2a, this represents the length of the sequence of target polynucleotide analyzed. Figure 2b illustrates that in Example 2b, 100.0% of the 10 target polynucleotides of the simulation were again extended by at least 25 incorporated nucleotides. This illustrates that using a cycle period of 0.8 half-

- 30 -

lives, and repeating the cycles 60 times, allowed analysis of a 25 base sequence of 10 target polynucleotides.

[0105]　　　　Comparing the two simulations, it will be appreciated by those in the art that the use of short-cycles of sequencing overcame issues of reading long repeats of homopolymer stretches in sequencing by synthesis, without using blocking moieties, as only a few nucleotides were incorporated per cycle. Comparing Examples 2a and 2b, the long cycles in 2a resulted in 40% of the extended complementary strands having two or less homopolymer nucleotide incorporations per cycle. Conversely, the short cycles in 11b resulted in 90% of the extended complementary strands having two or less homopolymer nucleotide incorporations per cycle, facilitating quantification. That is, as explained more thoroughly above, shorter reads can be quantitated to determine the number of nucleotides incorporated, for example, where the nucleotides are of the same

[0106]　　　　Comparing Examples 2a and 2b also indicated that a greater number of repeated cycles were needed to analyze a given length of sequence when using shorter cycles. That is, the 10 half-lives cycle was repeated 12 times to result in 100.0% of the 10 complementary strands being extended by at least 25 nucleotides, whereas the 0.8 half-lives cycle was repeated 60 times to obtain this same result and thereby analyze the 25 nucleotides sequence.

[0107]　　　・　　Nonetheless, many aspects of the repeated cycles may be automated, for example, using micro fluidics for washing nucleotides to sites of anchored target polynucleotides, and washing out unincorporated nucleotides to halt each cycle.

[0108]　　　　As discussed herein, below is a copy of the source code for the simulation of short-cycle sequencing.

Source Code for Simulation of Short Cycle Sequencing

```
Option Explicit 'all variables must be declared
Option Base 1   'array pointers start at '1' not '0'



'-------Constant Declarations--------------------------------
Const NoColor = 0
Const Black = 1
```

```
        Const White = 2
        Const Red = 3
        Const Green = 4
        Const Blue = 5
5       Const Yellow = 6
        Const Magenta = 7
        Const Cyan = 8
        Const A = Red
        Const G = Green
10      Const T = Blue
'       Const C = Yellow
        Const TENTH_HL = 0.93305



15      '--------Variable Declarations-------------------------------
        'Note: HL is short for half-life
        Dim MaxHalfLives As Integer 'The maximum number of half-lives the experiment will be
        run X10 for each wash cycle
        Dim HalfLives 'the Half Life variable is stepped in increments 0.1 half lives during every
20      wash cycle until the max is reached


        Dim N, I, J, K, L, X, Y, Temp As Integer
        Dim WashCyclesMax, WashCycles 'A wash cycle is completed after flowing each of
        AGT&C
25      Dim Molecule, Base, BaseType, Position As Integer
        Dim TempReal As Single
        Dim RandomMoleculesMax
        Dim HomoPolymersMax
        Dim MoleculesMax As Integer
30


        '------------the following three variables used to slow things down for second base
        Dim Longer_HL As Single
```
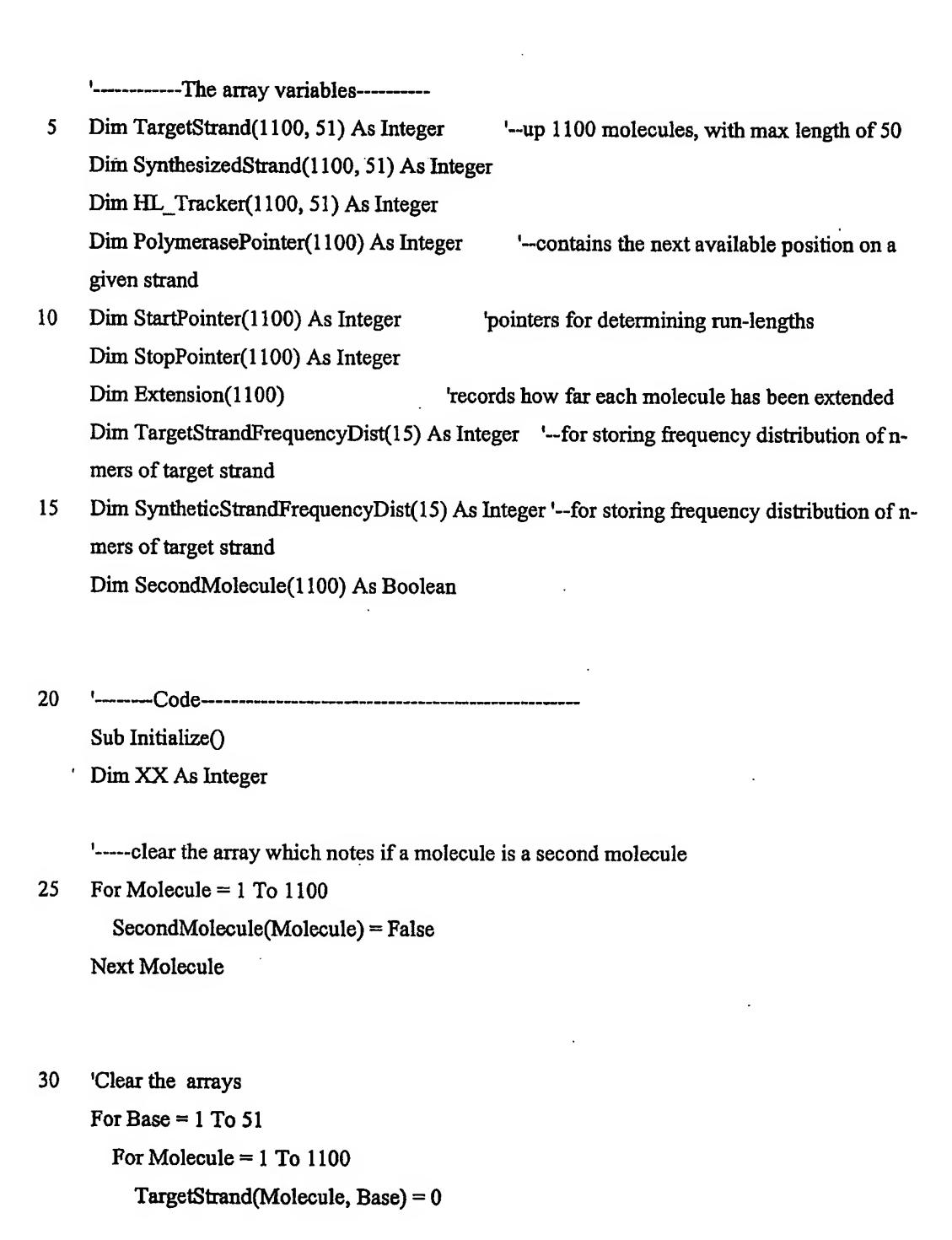
```
Dim SecondMoleculeFactor As Integer


        '-----------The array variables----------
5       Dim TargetStrand(1100, 51) As Integer          '--up 1100 molecules, with max length of 50
        Dim SynthesizedStrand(1100, 51) As Integer
        Dim HL_Tracker(1100, 51) As Integer
        Dim PolymerasePointer(1100) As Integer         '--contains the next available position on a
        given strand
10      Dim StartPointer(1100) As Integer              'pointers for determining run-lengths
        Dim StopPointer(1100) As Integer
        Dim Extension(1100)                            'records how far each molecule has been extended
        Dim TargetStrandFrequencyDist(15) As Integer   '--for storing frequency distribution of n-
        mers of target strand
15      Dim SyntheticStrandFrequencyDist(15) As Integer '--for storing frequency distribution of n-
        mers of target strand
        Dim SecondMolecule(1100) As Boolean



20      '--------Code----------------------------------------------------
        Sub Initialize()
        Dim XX As Integer


        '-----clear the array which notes if a molecule is a second molecule
25      For Molecule = 1 To 1100
            SecondMolecule(Molecule) = False
        Next Molecule



30      'Clear the arrays
        For Base = 1 To 51
            For Molecule = 1 To 1100
                TargetStrand(Molecule, Base) = 0
```

```
        SynthesizedStrand(Molecule, Base) = 0
        HL_Tracker(Molecule, Base) = 0
        PolymerasePointer(Molecule) = 1
    Next Molecule
5   Next Base


    For XX = 1 To 15  '--clear the frequency distribution list
        TargetStrandFrequencyDist(XX) = 0
        SyntheticStrandFrequencyDist(XX) = 0
10  Next XX


    For XX = 1 To 9
    Worksheets("In Out").Cells(5 + XX, 10).Value = ""
    Next XX

15


    With Worksheets("In Out")
        'Get the "front panel" input values
        TempReal = .Range("D4").Value
20      MaxHalfLives = Int(TempReal * 10)
        WashCyclesMax = .Range("D7").Value
        RandomMoleculesMax = .Range("D9").Value
        If RandomMoleculesMax > 1000 Then RandomMoleculesMax = 1000
        HomoPolymersMax = .Range("D11").Value
25      If HomoPolymersMax > 100 Then HomoPolymersMax = 100
        MoleculesMax = RandomMoleculesMax + HomoPolymersMax
        SecondMoleculeFactor = .Range("D14").Value
        Longer_HL = Exp(-0.0693 / SecondMoleculeFactor)


30

        '----Clear the output values
        .Range("D20").Value = ""
        .Range("D21").Value = ""
```

```vbnet
        .Range("D22").Value = ""
        .Range("E24").Value = ""
        .Range("E25").Value = ""
        .Range("E26").Value = ""
    End With
    'Worksheets("In Out").Range("E2").Value = Longer_HL   'Display the Longer_HL value
    'Clear the output area & Fill Row headings
    With Worksheets("Molecules")
        .Range("B2:AY4006").ClearContents
        .Range("B2:AY4006").Interior.ColorIndex = NoColor
        For XX = 1 To 1100
            .Cells(3 + XX * 4, 1).Value = XX 'Add the row headings as running numbers
        Next XX
        .Range("B3").Value = "Current Wash Cycle is:"
        .Range("L3").Value = "Current 'Half-Life' is:"
        .Range("U3").Value = "Current Base in the reaction is:"


    End With


    Randomize '---Seed the Random Number Generator


 End Sub


 Sub DrawSynthesizedStrands()
 Dim TempMolecule, TempBase As Integer


 With Worksheets("molecules")
 For TempBase = 1 To 50
     For TempMolecule = 1 To MoleculesMax
         If SynthesizedStrand(TempMolecule, TempBase) = Blue Then
             .Cells(TempMolecule * 4 + 2, TempBase + 1).Font.ColorIndex = 2
         Else
             .Cells(TempMolecule * 4 + 2, TempBase + 1).Font.ColorIndex = 0
```

```
        End If
        .Cells(TempMolecule * 4 + 2, TempBase + 1).Interior.ColorIndex =
SynthesizedStrand(TempMolecule, TempBase)
            If HL_Tracker(TempMolecule, TempBase) > 0 Then
5               .Cells(TempMolecule * 4 + 2, TempBase + 1).Value = HL_Tracker(TempMolecule,
TempBase)
            End If
        Next TempMolecule
    Next TempBase
10  End With


    End Sub



15  Sub CreateTargetStrands()
    Dim TempRand As Integer


    For Base = 1 To 50
        For Molecule = 1 To RandomMoleculesMax
20          TempRand = Int(4 * Rnd + 3) 'random number of value 3,4,5 or 6
            'If TempRand = Blue Then TempRand = Cyan 'turn blue into cyan
            TargetStrand(Molecule, Base) = TempRand
            Worksheets("Molecules").Cells(Molecule * 4 + 3, Base + 1).Interior.ColorIndex =
TargetStrand(Molecule, Base)
25      Next Molecule
    Next Base



    '--now draw molecules with long stretches of homopolymers
30  For Base = 1 To 50
        For Molecule = RandomMoleculesMax + 1 To MoleculesMax
            TargetStrand(Molecule, Base) = A
```

```
        Worksheets("Molecules").Cells(Molecule * 4 + 3, Base + 1).Interior.ColorIndex =
TargetStrand(Molecule, Base)
        Next Molecule
    Next Base

5

    End Sub


    Sub Synthesize()
10  Dim MoleculeSynthesized As Integer
    Dim TempPointer As Integer
    Dim Parameter As Single


15  For Molecule = 1 To 1100   'clear array which shows if molecule is a second molecule
        SecondMolecule(Molecule) = False
    Next Molecule


    For BaseType = A To C  'Cover each of AGT&C
20      If BaseType = A Then Worksheets("Molecules").Range("AD3").Value = "A"
        If BaseType = G Then Worksheets("Molecules").Range("AD3").Value = "G"
        If BaseType = T Then Worksheets("Molecules").Range("AD3").Value = "T"
        If BaseType = C Then Worksheets("Molecules").Range("AD3").Value = "C"


25


    For HalfLives = 1 To MaxHalfLives
    Worksheets("Molecules").Range("R3").Value = HalfLives / 10
        For Molecule = 1 To MoleculesMax
30          If SecondMolecule(Molecule) = False Then Parameter = TENTH_HL Else Parameter
= Longer_HL


        '-----------If we're flowing in A's, we attempt to polymerize only to T's
```

```
        If BaseType = A And TargetStrand(Molecule, PolymerasePointer(Molecule)) = T
Then

            If Rnd > Parameter Then MoleculeSynthesized = 1 Else MoleculeSynthesized = 0
'did molecule go?
5           If MoleculeSynthesized = 1 Then
                SecondMolecule(Molecule) = True
                SynthesizedStrand(Molecule, PolymerasePointer(Molecule)) = A
                HL_Tracker(Molecule, PolymerasePointer(Molecule)) = WashCycles
                PolymerasePointer(Molecule) = PolymerasePointer(Molecule) + 1
10              If PolymerasePointer(Molecule) > 50 Then PolymerasePointer(Molecule) = 50
            End If
        End If


        '------------If we're flowing in T's, we attempt to polymerize only to A's
15      If BaseType = T And TargetStrand(Molecule, PolymerasePointer(Molecule)) = A
Then

            If Rnd > Parameter Then MoleculeSynthesized = 1 Else MoleculeSynthesized = 0
'did molecule go?
            If MoleculeSynthesized = 1 Then
20              SecondMolecule(Molecule) = True
                SynthesizedStrand(Molecule, PolymerasePointer(Molecule)) = T
                HL_Tracker(Molecule, PolymerasePointer(Molecule)) = WashCycles
                PolymerasePointer(Molecule) = PolymerasePointer(Molecule) + 1
                If PolymerasePointer(Molecule) > 50 Then PolymerasePointer(Molecule) = 50
25          End If
        End If


        '----------If we're flowing in G's, we attempt to polymerize only to C's
        If BaseType = G And TargetStrand(Molecule, PolymerasePointer(Molecule)) = C
30 Then

            If Rnd > Parameter Then MoleculeSynthesized = 1 Else MoleculeSynthesized = 0
'did molecule go?
            If MoleculeSynthesized = 1 Then
```

```
              SecondMolecule(Molecule) = True

              SynthesizedStrand(Molecule, PolymerasePointer(Molecule)) = G

              HL_Tracker(Molecule, PolymerasePointer(Molecule)) = WashCycles

              PolymerasePointer(Molecule) = PolymerasePointer(Molecule) + 1

5             If PolymerasePointer(Molecule) > 50 Then PolymerasePointer(Molecule) = 50
          End If
      End If


      '-----------If we're flowing in C's, we attempt to polymerize only to G's

10    If BaseType = C And TargetStrand(Molecule, PolymerasePointer(Molecule)) = G
  Then
          If Rnd > Parameter Then MoleculeSynthesized = 1 Else MoleculeSynthesized = 0
  'did molecule go?
          If MoleculeSynthesized = 1 Then

15            SecondMolecule(Molecule) = True

              SynthesizedStrand(Molecule, PolymerasePointer(Molecule)) = C

              HL_Tracker(Molecule, PolymerasePointer(Molecule)) = WashCycles

              PolymerasePointer(Molecule) = PolymerasePointer(Molecule) + 1

              If PolymerasePointer(Molecule) > 50 Then PolymerasePointer(Molecule) = 50
20        End If
      End If


      Next Molecule
      'DrawSynthesizedStrands '--for now, display is refreshed after each increment of half
25 life for a given base
      Next HalfLives


  Next BaseType


30 End Sub
```

'---Develop an analysis of the distribution of homopolymers in the full-length targets, report as a frequency

'---distribution of n-mers


```
5    Sub AnalyzeTargetStrands()
     Dim CurrentBase As Integer
     Dim BasesAhead As Integer
     Dim N As Integer
     Dim NumberedBases(50) As Integer
10   Dim RunLengths(50) As Integer



     For N = 1 To 15  '--clear the frequency distribution list
         TargetStrandFrequencyDist(N) = 0      '
15       SyntheticStrandFrequencyDist(N) = 0
     Next N


     For Molecule = 1 To MoleculesMax 'Identify Changes among bases
         NumberedBases(1) = 1
20       'Worksheets("Molecules").Cells(4 + Molecule * 4, 2).Value = NumberedBases(1) 'take this
     out. For display only
         For Base = 2 To 50
             If TargetStrand(Molecule, Base - 1) <> TargetStrand(Molecule, Base) Then
                 NumberedBases(Base) = 1
25           Else
                 NumberedBases(Base) = 0
             End If
             'Worksheets("Molecules").Cells(4 + Molecule * 4, Base + 1).Value =
     NumberedBases(Base) 'take this out. For display only
30       Next Base


         '-------------compute run lengths
           """But first we've got a boundary condition problem for the first base--we solve it here!!
```

```
RunLengths(1) = 1
'Worksheets("Molecules").Cells(5 + Molecule * 4, 2).Value = RunLengths(1)


For Base = 2 To 50
    If NumberedBases(Base) = 1 Then
        RunLengths(Base) = 1
    Else
        RunLengths(Base) = RunLengths(Base - 1) + 1
    End If
    'Worksheets("Molecules").Cells(5 + Molecule * 4, Base + 1).Value = RunLengths(Base)
Next Base


'----save only the terminal value of a run length
For Base = 1 To 49
    If RunLengths(Base + 1) > RunLengths(Base) Then RunLengths(Base) = 0
    'Worksheets("Molecules").Cells(6 + Molecule * 4, Base + 1).Value = RunLengths(Base)
Next Base
    'Worksheets("Molecules").Cells(6 + Molecule * 4, 50 + 1).Value = RunLengths(50)
'boundary condition



'----Now determine the frequency distribution of each N-mer
For Base = 1 To 50
    If RunLengths(Base) = 1 Then TargetStrandFrequencyDist(1) =
TargetStrandFrequencyDist(1) + 1
    If RunLengths(Base) = 2 Then TargetStrandFrequencyDist(2) =
TargetStrandFrequencyDist(2) + 1
    If RunLengths(Base) = 3 Then TargetStrandFrequencyDist(3) =
TargetStrandFrequencyDist(3) + 1
    If RunLengths(Base) = 4 Then TargetStrandFrequencyDist(4) =
TargetStrandFrequencyDist(4) + 1
    If RunLengths(Base) = 5 Then TargetStrandFrequencyDist(5) =
TargetStrandFrequencyDist(5) + 1
```

```
        If RunLengths(Base) = 6 Then TargetStrandFrequencyDist(6) =
TargetStrandFrequencyDist(6) + 1
        If RunLengths(Base) = 7 Then TargetStrandFrequencyDist(7) =
TargetStrandFrequencyDist(7) + 1
5       If RunLengths(Base) = 8 Then TargetStrandFrequencyDist(8) =
TargetStrandFrequencyDist(8) + 1
        If RunLengths(Base) >= 9 Then TargetStrandFrequencyDist(9) =
TargetStrandFrequencyDist(9) + 1
        Next Base
10  Next Molecule
    For I = 1 To 9
    Worksheets("In Out").Cells(5 + I, 10).Value = TargetStrandFrequencyDist(I) 'copy to the
spreadsheet
    Next I
15
    End Sub



    Sub AnalyzeResults()
20  Dim N As Integer
    Dim TwentyFiveMer, TwentyFiveMerAccumulator As Integer
    Dim LongestLength, ShortestLength  As Integer
    Dim TempSum, Min, Max As Integer
    Dim AverageLength As Single
25
    '----First we analyze the data about the degree of extension
    For N = 1 To 1100          "clear the extension array.
      Extension(N) = 0
    Next N
30


    For Molecule = 1 To MoleculesMax
      N = 0
```

```
For Base = 1 To 50

    If SynthesizedStrand(Molecule, Base) <> 0 Then N = Base

    'N = 1 'debug statement

    Next Base
```

5

```
    Extension(Molecule) = N

    'Worksheets("In Out").Range("C13").Value = Extension(N) 'debug statement

Next Molecule "---we now have an array of maximum lengths of each strand in Extension.

We can now compute...
```

10

```
'First we do the average:

TempSum = 0

For N = 1 To 1100

    TempSum = Extension(N) + TempSum '--grand total
```

15

```
Next N

AverageLength = TempSum / MoleculesMax

Worksheets("In Out").Range("D22").Value = AverageLength
```

20

```
'Now we find the Min and Max

Max = 0

Min = 50

For N = 1 To MoleculesMax

    If Max > Extension(N) Then Max = Max Else Max = Extension(N)
```

25

```
    If Min < Extension(N) Then Min = Min Else Min = Extension(N)

Next N

Worksheets("In Out").Range("D20").Value = Max

Worksheets("In Out").Range("D21").Value = Min
```

30

```
'Determine what fraction of molecules are more than 25 bases long

TwentyFiveMerAccumulator = 0

For N = 1 To MoleculesMax
```

```
        If Extension(N) > 24 Then TwentyFiveMerAccumulator = TwentyFiveMerAccumulator +
    1
    Next N
    Worksheets("In Out").Range("E26").Value = TwentyFiveMerAccumulator / MoleculesMax

5
    End Sub



    Sub AnalyzeSynthesizedStrands()

10
    Dim CurrentBase As Integer
    Dim BasesAhead As Integer
    Dim N As Integer
    Dim NumberedBases(51) As Integer
15  Dim RunLengths(51) As Integer
    Dim TwoHitAccumulator, ThreePlusHitAccumulator, TwoHit, ThreeHit As Integer



    TwoHitAccumulator = 0
20  ThreePlusHitAccumulator = 0



    For I = 1 To 50
        NumberedBases(I) = 3
25  Next I


    For Molecule = 1 To MoleculesMax 'Identify Changes among bases
        NumberedBases(1) = 1
        'Worksheets("Molecules").Cells(1 + Molecule * 4, 2).Value = NumberedBases(1) 'take this
30  out. For display only
            For Base = 2 To Extension(Molecule)
                If SynthesizedStrand(Molecule, Base - 1) <> SynthesizedStrand(Molecule, Base) Or
    HL_Tracker(Molecule, Base - 1) <> HL_Tracker(Molecule, Base) Then
```

```
            NumberedBases(Base) = 1
        Else
            NumberedBases(Base) = 0
        End If
5       'Worksheets("Molecules").Cells(1 + Molecule * 4, Base + 1).Value =
    NumberedBases(Base) 'take this out. For display only
        Next Base


        '--------------compute run lengths
10      ""'But first we've got a boundary condition problem for the first base--we solve it here!!
        RunLengths(1) = 1
        'Worksheets("Molecules").Cells(1 + Molecule * 4, 2).Value = RunLengths(1)


        For Base = 2 To Extension(Molecule)
15          If NumberedBases(Base) = 1 Then
                RunLengths(Base) = 1
            Else
                RunLengths(Base) = RunLengths(Base - 1) + 1
            End If
20          'Worksheets("Molecules").Cells(1 + Molecule * 4, Base + 1).Value = RunLengths(Base)
        Next Base


        '----save only the terminal value of a run length
        For Base = 1 To Extension(Molecule)
25          If RunLengths(Base + 1) > RunLengths(Base) Then RunLengths(Base) = 0
                'Worksheets("Molecules").Cells(1 + Molecule * 4, Base + 1).Value = RunLengths(Base)
        Next Base
            'Worksheets("Molecules").Cells(1 + Molecule * 4, 50 + 1).Value =
    RunLengths(Molecule) 'boundary condition
30
        TwoHit = 0
        ThreeHit = 0
        For Base = 1 To Extension(Molecule)
```

```
If RunLengths(Base) = 2 Then
    Worksheets("Molecules").Cells(1 + Molecule * 4, Base + 1).Interior.ColorIndex =
Magenta
    TwoHit = 1
End If


If RunLengths(Base) > 2 Then
    Worksheets("Molecules").Cells(1 + Molecule * 4, Base + 1).Interior.ColorIndex =
Cyan
    ThreeHit = 1
End If
Next Base
'--Now determine what fraction of molecules have either 2 bases or 3+ base hits and report
results
TwoHitAccumulator = TwoHitAccumulator + TwoHit
ThreePlusHitAccumulator = ThreePlusHitAccumulator + ThreeHit
Next Molecule


Worksheets("In Out").Range("E24").Value = TwoHitAccumulator / MoleculesMax
Worksheets("In Out").Range("E25").Value = ThreePlusHitAccumulator / MoleculesMax


End Sub


Public Sub Main_Line()
    Initialize


'--Creates the new strands based on number of washes for varying degrees of completion
per cycle
If MoleculesMax > 0 And WashCyclesMax > 0 Then
    CreateTargetStrands
    AnalyzeTargetStrands
    For WashCycles = 1 To WashCyclesMax 'Do the desired number of wash cycles
```

- 46 -

```
Worksheets("Molecules").Range("I3").Value = WashCycles
Synthesize
Next WashCycles
DrawSynthesizedStrands
AnalyzeResults
AnalyzeSynthesizedStrands
End If
End Sub
```

Example 3

[0109]   Figure 2 illustrates yet another simulated analysis of a number of target polynucleotides u sing s hort-cycle sequencing.   The s imulation w as run using the program described in Examples 2a and 2b but using a larger number of target polynucleotides.

[0110]   That is, in this simulation, the input values used were a cycle period of 0.8 half-lives, 60 repeats of the cycle, and 200 target polynucleotide strands. Figure 2 illustrates the results obtained.   Homopolymers stretches which occurred   in the same simulated complementary strand are highlighted in magenta wherever nucleotides of the same base type were incorporated in a row, and in cyan wherever more than two nucleotides of the same base type were incorporated in a row.

[0111]   The output values obtained were 48 incorporations in the longest extended complementary strand, 20 in the shortest, and 32.00 as the average number of incorporations for the 200 simulatedly extended complementary strands.

[0112]   Further, the percentage of growing stands extended by two or more nucleotides in a homopolymer stretch was 78.5%; and the percentage of growing strands extended by three or more nucleotides in a homopolymer stretch was 4.0%.   That is, using a cycle period of 0.8 half-lives resulted in 96.0% of the complementary strands being extended by two or less nucleotides in a homopolymer stretch per cycle of incorporation.   Moreover, 95.5% of the 200 target polynucleotides of the simulation were extended by at least 25 incorporated nucleotides, while 100% were extended by at least 20 nucleotides.   This illustrated that using a cycle period of 0.8 half-lives, and repeating the cycles 60 times, allows analysis of a 20 base sequence of 200 target polynucleotides.

Example 4

[0113]   This example demonstrates a method according to the invention in which a single nucleotide in a position in a nucleic acid sequence is identified.   A template-bound

- 47 -

primer is sequentially exposed first to a labeled nucleotide and then to an unlabeled
nucleotide of the same type under conditions and in the presence of reagents that allow
template-dependent primer extension. The template is analyzed in order to determine
whether the first nucleotide is incorporated in the primer at the first position or not. If not,

5   then the sequential exposure to labeled and unlabeled nucleotides is repeated using another
type of nucleotide until one such nucleotide is determined to have incorporated at the first
position. Once an incorporated nucleotide is determined, the identity of the nucleotide in the
position in the nucleic acid sequence is identified as the complementary nucleotide.

Example 5

10   [0114]   In this example, a series of reactions are performed as described above in
Example 1. A nucleic acid primer is hybridized to a target nucleic acid at a primer binding
site in the target. The primer comprises a donor fluorophore. The hybridized primer is
exposed to a first nucleotide comprising an acceptor fluorophore comprising a blocking
moiety that, when incorporated into the primer, prevents further polymerization of the primer.

15   The presence or absence of fluorescent emission from each of the donor and the acceptor is
determined. A nucleotide that has been incorporated into the primer via complementary base
pairing with the target is identified by the presence of fluorescent emission from the acceptor,
and a sequence placeholder is identified as the absence of fluorescent emission from the
donor and the acceptor. A sequence of the target nucleic acid is complied based upon the

20   sequence of the incorporated nucleotides and the placeholders.

· Claims

1   We claim:

2   1.   A method for sequencing a nucleic acid template, the method comprising the steps of:

3   (a)   exposing a nucleic acid template to a primer capable of hybridizing to said

4   template and a polymerase capable of catalyzing nucleotide addition to said primer;

5   (b)   adding a labeled nucleotide for a predetermined time, said predetermined time

6   being coordinated with an amount of polymerization inhibition such that on average

7   only 0, 1, or 2 labeled nucleotides are added to said primer;

8   (c)   removing excess labeled nucleotide;

9   (d)   neutralizing label in any incorporated nucleotide;

10   (e)   repeating steps a, b, c, and d at least once; and

11   (f)   determining a sequence of said template based upon the order of incorporation

12   of said labeled nucleotides.

1   2.   A method for conducting a nucleic acid sequencing reaction, the method comprising

2   the steps of:

3   providing a nucleic acid template and a primer capable of hybridizing to a

4   portion of said template, thereby to form a primed template;

5   exposing said primed template to a nucleotide for a period of time that is

6   statistically insufficient for incorporation of more nucleotides than are resolvable by a

7   detection system used to detect incorporation of said nucleotide into said primer;

8   detecting incorporation of said nucleotide;

9   neutralizing label in an incorporated nucleotide;

10   repeating said providing, exposing, detecting, and neutralizing steps at least

11   once; and

12   determining a sequence of said template based upon the order of nucleotides

13   incorporated into said primer.

1   3.   A method for identifying a nucleotide incorporated into a primer in template-

2   dependent nucleic acid sequencing, the method comprising the steps of:

3   conducting a plurality of base incorporation cycles, wherein each cycle

4   comprises exposing a template nucleic acid to a labeled nucleotide that is not a chain-

5   terminating nucleotide, wherein said labeled nucleotide is incorporated into a primer

6   hybridized to said template if said nucleotide is capable of hybridizing to a template

7   nucleotide immediately upstream of said primer, and wherein there is about a 99%

8   probability that two or fewer of said nucleotides are incorporated into the same primer

9   strand per cycle; and

10   identifying incorporated nucleotides.

1   4.   A method for template-dependent nucleic acid sequencing, the method comprising the

2   steps of:

3   (a)   exposing a template nucleic acid to a labeled nucleotide under conditions that

4   allow incorporation of said nucleotide into a primer attached to said template;

5   (b)   removing unhybridized nucleotide from said template at a time after said

6   exposing step that is sufficient for incorporation of no more than about two of said

7   nucleotides per template;

8   (c)   determining if a nucleotide is incorporated into said primer;

9   (d)   identifying any incorporated nucleotide;

10   (e)   repeating steps a, b, c, and d; and

11   (f)   compiling a sequence of said template based upon the sequence of nucleotides

12   incorporated into said primer.

1   5.   A method for template-dependent nucleic acid sequencing, the method comprising the

2   steps of:

3   (a)   exposing a template nucleic acid to a labeled nucleotide under conditions that

4   allow incorporation of said nucleotide into a primer attached to said template;

5   (b)   removing unhybridized nucleotide at a time after said exposing step that is

6   statistically insufficient for incorporation of a greater number of nucleotides than are

7   individually optically resolvable during a predetermined detection period;

8   (c) detecting incorporation of individual labeled nucleotides during said detection

9   period;

10   (d)   neutralizing label present in incorporated nucleotides;

11   (e)   repeating steps a, b, c, and d at least once; and

12   (f)   compiling a sequence of said template based upon an order of incorporated

13   nucleotides.

1   6.   A method for nucleic acid sequencing, the method comprising the steps of:

2   (a)   selecting a nucleic acid template to be sequenced;

3   (b)   exposing said template to a primer that is capable of hybridizing to a portion

4   of said template to form a primed template;

5   (c)   selecting a desired number of nucleotides to be added to said primer;

6   (d)   determining a reduction in the rate at which a second nucleotide is added to

7   said primer given that a first labeled nucleotide has already been added to said primer;

8   (e)   identifying a tolerable rate of erroneous detection of an incorporated

9   nucleotide;

10   (f)   exposing said primed template to a labeled nucleotide

11   (g)   removing unincorporated labeled nucleotide at a time after said exposing step

12   that is determined based upon said desired number, said rate reduction, and said

13   tolerable error, such that said time is statistically insufficient for incorporation of

14   more nucleotides than are resolvable by a detection system used to detect

15   incorporation of said nucleotide into said primer;

16   (h)   identifying incorporated nucleotide;

17   (i)   neutralizing label present in said incorporated nucleotide;

18   (j)   repeating steps f, g, and h at least once; and

19   (k)   determining a sequence of said template based upon an order of said

20   incorporated nucleotides.

1   7.   A method for sequencing a template nucleic acid, the method comprising the steps of:

2   (a)   conducting a cycle of template-dependent nucleic acid primer extension in the

3   presence of a polymerase and a labeled nucleotide;

4   (b)   inhibiting polymerase activity such that it is statistically unlikely that more

5   than about 2 nucleotides are incorporated into the same primer strand in said cycle;

6   (c)   washing unincorporated labeled nucleotide away from said template;

7   (d)   detecting any incorporation of said labeled nucleotide;

8   (e)   neutralizing label in any incorporated labeled nucleotide;

9   (f)   removing said inhibition;

10   (g)   repeating steps a, b, c, d, e, and f; and

11   (h)   compiling a sequence of said template based upon the sequence of nucleotides

12   incorporated into said primer.

1   8.   A method for sequencing a target nucleic acid, the method comprising the steps of:

2   conducting a plurality of primer extension cycles, wherein each cycle

3   comprises the steps of

4   exposing a target nucleic acid to a primer capable of hybridizing to said target

5   thereby to form a primed target,

6   exposing said primed target to a labeled nucleotide in the presence of a nucleic

7   acid polymerase,

8   coordinating transient inhibition of said polymerase and time of exposure to

9   said labeled nucleotide such that it is statistically likely that at least one of said

10   labeled nucleotide is incorporated in said primer, but statistically unlikely that more

11   than two of said labeled nucleotide are incorporated in said primer.

1   9.   A method for identifying a nucleotide incorporated into a primer in a template-

2   dependent primer extension reaction, the method comprising the steps of:

3   exposing a template nucleic acid to a primer capable of hybridizing to said

4   template and a polymerase capable of catalyzing template-dependent nucleotide

5   addition to said primer;

6   adding a labeled nucleotide;

7   optically detecting whether said labeled nucleotide is incorporated into said

8   primer, wherein said detecting occurs at a rate sufficient to detect 1, but no more than

9   2, incorporated nucleotides per detection cycle; and

10   identifying an incorporated nucleotide.

1   10.   A method for determining the sequence of a template nucleic acid, the method

2   comprising the steps of:

3   (a)   exposing a nucleic acid template to a primer capable of hybridizing to a

4   portion of said template in order to form a template/primer complex reaction mixture;

5   (b)   adding a labeled nucleotide in the presence of a polymerase to said mixture

6   under conditions that promote incorporation of said nucleotide into said primer if said

7   nucleotide is complementary to a nucleotide in said template that is downstream of

8   said primer;

9   (c)   coordinating removal of said labeled nucleotide and inhibition of said

10   polymerase so that no more than about 2 nucleotides are incorporated into the same

11   primer;

12   (d)   identifying labeled nucleotide that has been incorporated into said primer;

13   (e)   repeating steps a, b, c, and d at least once; and

14   (f)   determining a sequence of said template based upon the order of said

15   nucleotides incorporated into said primer.

1   11.   A method for identifying a nucleotide present in a template sequence, the method

2   comprising the steps of:

3   exposing a template nucleic acid to a primer capable of hybridizing to a

4   portion of said template upstream of a region of said template to be sequenced;

5   introducing a labeled nucleic acid and a polymerase to said template under

6   conditions wherein said labeled nucleic acid will be incorporated in said primer if said

7   labeled nucleic acid is capable of hybridizing with a base downstream of said primer;

8   and

9   controlling the rate of said incorporation by limiting the time of exposure of

10   said labeled nucleic acid to said template or by inhibiting said polymerase at a

11   predefined time after exposure of said template to said labeled nucleotide;

12   detecting incorporation of said labeled nucleotide into said primer; and

13   identifying said nucleotide in said template as the complement of labeled

14   nucleotide incorporated into said primer.

1   12.   A method for sequencing a target nucleic acid, the method comprising the steps of:

2   hybridizing a nucleic acid primer comprising a donor fluorophore to a target

3   nucleic acid at a primer binding site in said target;

4   exposing said hybridized primer to a first nucleotide comprising an acceptor

5   fluorophore that, when incorporated into said primer, prevents further polymerization

6   of said primer;

7   detecting the presence or absence of fluorescent emission from each of said

8   donor and said acceptor;

9   identifying a nucleotide that has been incorporated into said primer via

10   complementary base pairing with said target as the presence of fluorescent emission

11   from said acceptor;

12   identifying a sequence placeholder as the absence of fluorescent emission

13   from said donor and said acceptor; and

14   repeating said exposing, detecting, and each of said identifying steps, thereby

15   to compile a sequence of said target nucleic acid based upon the sequence of said

16   incorporated nucleotides and said placeholders.

1   13.   A method for identifying a placeholder in a nucleic acid sequence determined by

2   synthesis, the method comprising the steps of:

3   hybridizing a nucleic acid primer comprising a donor fluorophore to a target

4   nucleic acid at a primer binding site in said target;

5   exposing said hybridized primer to a first nucleotide comprising an acceptor

6   fluorophore that, when incorporated into said primer, prevents further polymerization

7   of said primer;

8   determining whether there is fluorescent emission from said donor and said

9   acceptor; and

10   identifying a placeholder in said nucleic acid sequence as the absence of

11   emission in both said donor and said acceptor.

1 14.   A method for sequencing a nucleic acid, the method comprising the steps of:

2   exposing a template-bound nucleic acid primer to a nucleotide comprising a

3   label that impedes progress of polymerase in the addition of a subsequent nucleotide;

4   determining whether said first labeled nucleotide has been incorporated into

5   said primer;

6   exposing said primer to an unlabeled first nucleotide if said first labeled

7   nucleotide has been incorporated into said primer;

8   repeating said exposing and determining steps with a second nucleotide if said

9   first nucleotide did not incorporate into said primer.

1 15.   The method of claim 2, further comprising

2   adding a first labeled nucleotide under conditions that optimize the

3   incorporation of one of said first nucleotide per primer strand;

4   removing unincorporated first labeled nucleotide;

5   detecting any incorporated first labeled nucleotide;

6   neutralizing label in said first labeled nucleotide; and

7   adding a second labeled nucleotide under conditions that optimize the

8   incorporation of one of said second nucleotides per primer strand.

9

1 16.   The method of claim 1 wherein said method does not utilize a blocking moiety.

1 17.   The method of claim 1 wherein said period of time is concluded by washing said

2 nucleotides not incorporated into said complementary strand.

1 18.   The method of claim 1 wherein said period of time is concluded by washing said

2 polymerization agent.

1 19.   The method of any of claim 1 wherein said period is no more than 5 half-lives of said

2 incorporation reactions.

1 20.   The method of claim 1 wherein said period is no more than 4 half-lives of said

2 incorporation reactions.

1 21.   The method of claim 1 wherein said period is no more than 3 half-lives of said

2 incorporation reactions.

1 22.   The method of claim 1 wherein said period is no more than 2 half-lives of said

2 incorporation reactions.

1 23.   The method of claim 1 wherein said period is no more than 1 half-lives of said

2 incorporation reactions.

1 24.   The method of claim 1 wherein said period is no more than 0.5 half-lives of said

2 incorporation reactions.

1 25.   The method of claim 1 wherein said period permits less than 5% chance of

2 incorporation of more than two of said nucleotides into said complementary strand.

1 26.   The method of claim 1 wherein said period is no more than 1 half-life of said

2 incorporation reactions and said wash cycles is repeated at least 40 times.

FIGURE 1

FIGURE 2A

FIGURE 2B

Target Strand Analysis

Length	Frequency
1-mer	5668
2-mer	1394
3-mer	333
4-mer	94
5-mer	27
6-mer	3
7-mer	1
8-mer	0
>=9-mer	1

Snythesized Strand Analysis

Length	Frequency
1-mer	
2-mer	
3-mer	
4-mer	
5-mer	
6-mer	
7-mer	
8-mer	
>=9-mer	

FIGURE 3

FIGURE 4

FIGURE 5

Image/Wash Cycle vs. dNTP T$_{1/2}$
(>80% Fract 2B+, <3% Frac. 3+ Homo)

FIGURE 6

Figure 7A

Figure 7B

5 U/mL Klenow Fragment at 37°C
Cy3 analogs for all